User-Centered Evaluation of a Virtual Environment Training System: Utility of User Perception Measures

Dawei Jia, Asim Bhatti, Chris Mawson, and Saeid Nahavandi

School of Engineering and Information Technology;
Centre for Intelligent System Research, Deakin University
dwj@deakin.edu.au

Abstract. This study assessed the utility of measures of Self-efficacy (SelfEfficacy) and Perceived VE efficacy (PVEefficacy) for quantifying how effective VEs are in procedural task training. SelfEfficacy and PVEefficacy have been identified as affective construct potentially underlying VE efficacy that is not evident from user task performance. The motivation for this study is to establish subjective measures of VE efficacy and investigate the relationship between PVEefficacy, SelfEfficacy and User task performance. Results demonstrated different levels of prior experience in manipulating 3D objects in gaming or computer environment (LOE3D) effects on task performance and user perception of VE efficacy. Regression analysis revealed LOE3D, SelfEfficacy, PVEefficacy explain significant portions of the variance in VE efficacy. Results of the study provide further evidence that task performance may share relationships with PVEefficacy and SelfEfficacy, and that affective constructs, such as PVEefficacy, and SelfEfficacy may serve as alternative, subjective measures of task performance that account for VE efficacy.

Keywords: User-based evaluation, Virtual Environment, Evaluation methodology.

1 Introduction

Immerging computing technologies, such as Virtual Reality (VR) is perceived to be effective in enhancing human abilities to complete complex tasks. A generic immersive VR system consists of a virtual environment (VE), advanced human-computer interface and models of interaction, these are useful for facilitating perception in such computer simulated 3D environments [18]. Enhanced perception is achieved through using displays that provide rich visual, auditory and haptics sensory information that allow human users to easily engage, immerse and interact with learning tasks [7]. A Virtual Environment (VE) is often used synonymously as VR to describe an environment based on real-world or abstract objects and data [17].

Effective design of a VE is often aimed at conveying to users the feeling of being “immersed”, “present”, “engaged”, “satisfied”, and or “enjoyed” in the simulated environment [9, 11, 12, 14, 16]. Moreover, to convey high level of self-efficacy- individual’s beliefs of his or her capability to organize and execute the behaviors to performing tasks successfully [19] are thought to be useful. It has been speculated that facilitation this sense of self-belief not only leads to higher level of acceptance and adaption of computer technology [3, 5] but also enhance task performance and outcomes [2].

R. Shumaker (Ed.): Virtual and Mixed Reality, LNCS 5622, pp. 196–205, 2009.

In the field of VE understanding how to use immersive technology to support the learning of abstract concepts or tasks and evaluating the degrees of effectiveness of how well a system is in assisting a user to achieve the intended learning, this present a substantial challenge for designer and evaluators of this technology. The challenges include users understanding, transfer of training and retention of trained techniques. Further, it is unlikely that a single evaluation factor or criteria construct will be capable of adequately assessing VE efficacy [10, 15, 17]. Many orientational, affective, cognitive and pedagogical issues are considered fundamental to VE efficacy [10]. However, there is currently no standard on the "best" way to quantify VE efficacy [17].

1.1 Measures of VE Efficacy in User-Based Evaluation

Typically, VE is measured objectively on user task performance. Common task performance measures used to evaluate the effectiveness of VE include time on task, speeds of completion and numbers of errors [12]. Additionally, having computer event driven recordings of all the experiments details, allowing for the incorporation of more accurate performance evaluation of the VE is also used widely in usability evaluation. Other objective measures are derived from physiological factors involve recording, such as heart-beat, blood pressure or eye movement over the course of the experiment. These are useful for ergonomic assessment of VE as they allow us to link physical responses directly to VE. Quantitative data produced from these objective measures is useful in showing "what" the users did, but they cannot be used to explain "how" or "why" user performed in a certain manner [18].

On the other hand, subjective measures, such as self-report in behavioral interview and questionnaires involve collecting both quantitative and qualitative data during a usability evaluation or user modeling, in which user behaviors are collected and assessed. Self-report data through behavioral questionnaire, for example is useful in collecting data of subjective views on particular aspect of interaction and learning experience with computer systems. Various multiple response modes such as physiological, motoric, and cognitive behaviors can be gathered using questionnaire [18]. It has advantages, such as an efficient use of time for both evaluator and respondent, and standardization of questions. In the field of VE, questionnaires are used quite frequently to elicit information about subjective phenomena [17]. Well designed questionnaire, such as Presence Questionnaire (PQ) by Witmer and Singer (1998) [20] have wide reaching effects and have been adapted extensively in evaluation of VE. More importantly, insight of user perception and preference of an interactive computer system can not be explained fully if only objective measures are used. For these reasons, there has been a call for subjective measures of VE in the literature.

1.2 Quantifying VE Efficacy through User Perception

Empirical evidence illustrates that perceptions such as self-efficacy (beliefs) and perceived efficacy of computer systems (attitude) can be influenced by the system design features in performance of cognitive or procedural tasks [5]. Self-efficacy is defined as an individual's expectancy in his or her capability to organize and execute the behaviors needed to successfully complete a task [1]. Perceived self-efficacy has been used to predict performance in decision making, cognitive task performance, and mathematical test scores [19], as well as proven to be beneficial in increase in problem-solving efficiency [6]. Prior research has also shown that perceived self-efficacy and attitudes toward computer are predictive of performance in computer mediated learning [13]. Perceived VE efficacy refers to user perception of how effective a VE is in assisting their interaction and learning experience, as well as learning outcomes. As an affective construct, perceived VE efficacy assesses VE quality from the users' point of view.

Subjective perception of VE and rating techniques has shown benefits in evaluating VE system [15]. In the field of usability engineering, user perception of computer technology plays an important role in evaluation. For example, users' perception of immersion, presence, engagement, satisfaction and enjoyment are associated with system design features. Draw on Fishbein and Ajzen's (1975) attitude paradigm from psychology, Davis (1993) [3] developed a technology acceptance model that addresses the beliefs (e.g. self-efficacy) and attitudes (e.g. perception) of the software systems on users' actual system usage, this plays a significant role in users' adoption of computer systems. For example, a person's belief about behaviour refers to his or her subjective likelihood that performing the behaviour will lead to a specified outcome; and attitude toward a behaviour is an affective evaluation of that behaviour. Because of the hypothesized benefit to performance, beliefs of self-efficacy and attitude towards computer system have been generally accepted as an evaluation criterion for computer mediated learning. On this point, perceived self-efficacy (SelfEfficacy) and perceived VE efficacy (PVEefficacy) are important in measuring performance in VE training system. We also hypothesize that if a system is effective, users should have higher perception of usability and learnability; and higher attention, comprehension, but lower cognitive load. In the field of VE, it is surprising to see a lack of work on incorporate self-efficacy and user perception measures to quantify VE efficacy. In line with other researchers [10, 17], we believe a reliable, repeatable and robust measure is needed to quantify VE efficacy.

The primary goal of this research is two fold: first, to determine if construct of the user perception measures of self-efficacy scale and perceived VE efficacy scale, can be used to quantify VE efficacy of an object assembly task; second, to explore the hypothesized relationships between self-efficacy beliefs, user perception of VE efficacy and task performance.

2 Hypotheses

Task performance and user perception are significantly affected by subjects' prior experience of manipulating 3D objects (LOE3D) in gaming or computer environments. Higher performance and perception will be associated with higher LOE3D. In addition, VE efficacy score will be significantly positive related to performance and perception on the object assembly task. VE efficacy also was expected to be significantly positive related to LOE3D.

3 Experiment

The validation of the proposed hypothesis is performed by training users in object assembly simulation called Virtual Training Environment (VTE) developed at Centre for Intelligent System Research (CISR), Deakin University. In addition, an empirical assessment of the object assembly simulation, based on the proposed evaluation framework [8], has been carried out. Thirty volunteers with different levels of experience in manipulating 3D objects in gaming or computing environment (LOE3D) performed a series of object assembly tasks in a virtual training system. The task involved selecting, rotating, releasing, inserting and manipulating 3D objects these tasks required users to utilize a data glove, a haptics device, a 3D mouse and a head-mounted display (HMD). Subjective assessments of SelfEfficacy, PVEefficacy were recorded along with objective assessment of task performance.

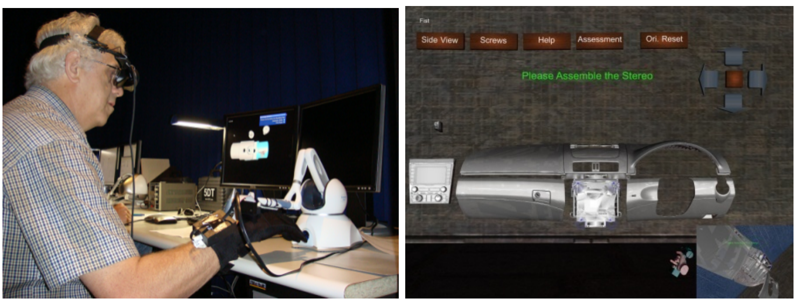

Fig. 1. Experiment setup (a) and training tasks (b)

Subjective measures on user perception of VE efficacy were captured through two questionnaires: Self-efficacy questionnaire (SEQ) or post-VE exposure questionnaire that measured self-efficacy (SelfEfficacy) and perceived VE efficacy questionnaire (PVEQ) or post-VE training test questionnaire that measured user perceived VE efficacy (PVEefficacy). Validation technique for questionnaire instrument [4] of factor analysis for data reduction of variables and Cronbach's Alpha for internal consistence were performed, which has shown the construct validity and reliability of these two measurement tools. Objective measure on task performance was captured through system logging file that automatically tracks user task performance and outcomes. A memory test was also conducted two weeks after the training test to assess users' long-term retention in the VE using a memory-test questionnaire (MTQ). Figure 2 represents the sequence of activities during the experiment. Upon entering the experimental environment, each subject was asked to complete a pre-test questionnaire (Pre-test Q). Each subject was then given a brief introduction of the system and performs a simple object assembly task, which serves as a pre-test of subject's ability to interact with, control and use various VE system control devices. SEQ was then filled out. Afterwards, a training test was presented to each subject, whom has 15 minutes to complete 7 object assembly tasks in the VE system. PVEQ was presented

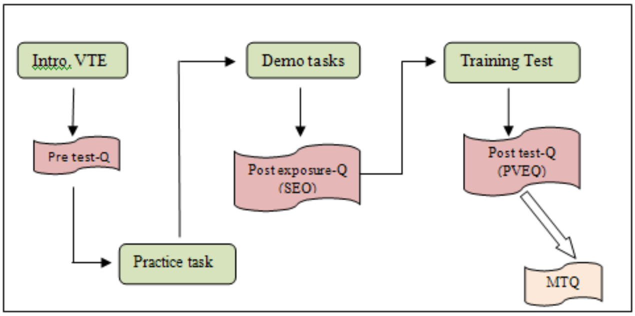

Fig. 2. Experiment Sequence

to the subject in the experimental environment. Lastly, an open-ended interview with each subject was carried out right after the test. Two weeks after the experimental test, subjects required to respond on the MTQ that requires them to recall their learning tasks or procedures in the VE training system.

4 Results and Discussion

VE efficacy was hypothesized to be significantly affected by different level of prior experience in manipulating 3D objects in gaming or computer environment (LOE3D). As VE efficacy was measured on TTS, SelfEfficacy, PVEefficacy and MMT, it was expected that people with higher level of LOE3D have higher self-efficacy beliefs, achieve better outome in training test, perceive the VE to be more effective and have higher achievement on the memory test.

4.1 Effect of Prior Experience

To assess the utility of prior experience for explaining task outcome, we used multiple predictors: computer use frequency (ComFreq), computer use history (CompHis), experience of manipulating 3D objects in gaming or computer environment (LOE3D), experience of manipulating 3D objects in VE environment (ExpVE). These were included in a multiple linear regression (MLR) model to predict training test score (TTS). Because of potential effects object assembly skills in real life may have influence on the subjects' performance in the VE, experience of using electronic tools for object assembly tasks (ExpTool), and perceived level of difficulty of assembly task (PdifTask) were included as predictors in this model. Finally, due to the potential effects of age and gender on training test score, and other response measures, these two variables were included in the model.

In general, the inclusion of these variables in the predictive model of training test score was aimed at avoiding biases in the parameter estimates; CompFreq, CompHis, LOE3D and ExpVE that might have occurred if variance due to prior object assembly skills (ExpTool, PdifTask) or individual differences were not taken into account However, it is anticipated that there were interrelationships among the variables. With this in mind, standard approach of multiple regression was performed, which allowed us to find out how the multiple predictors combine to influence the training test score. The regression model used to assess the utility of multiple predictors on training test score was structured as shown in equation 1.

$$\mathrm{TTS} = \beta_0 + \beta_1\mathrm{Age} + \beta_2\mathrm{Gender} + \beta_3\mathrm{CompFreq} + \beta_4\mathrm{CompHis} + \beta_5\mathrm{LOE3D} + \beta_6\mathrm{ExpVE} + \beta_7\mathrm{ExpTool} + \beta_8\mathrm{PdifTask} \tag{1}$$

Results of the standardized regression coefficients analysis indicated that this regression model predicts training test score well, F (2.404), $p<0.05$. Approximately 48% of the variability in training test score was explained by this model (R^2=0.478). The results also show that at the α=0.05 level, LOE3D is the most important predictor of training test score (*Beta*=0.567, p=0.032). More important, LOE3D alone, account for 38% of the variance of training test score, F=17.136, p=.000. Surprisingly, of the eight predictors, only subjects' prior experience of manipulating 3D object in gaming or computer environment contributes significantly (p=0.001) to the model. Correlation analysis (1-tailed) also confirms that LOE3D was significantly and positively correlated with training test score, r=.616, N=30, p=.000. In other words, people who are more experienced in manipulating 3D objects in gaming or computer environment tend to achieve higher training test score. In addition, a moderate but significant linear relationship between gender and training test score (r=0.321, N=30, $p=0.042$), and between ExpVE and training test score (r=.358, N=30, $p=.026$) were found. These results show that male tend to outperform than female, and people with more experience in manipulating 3D objects in VE achieved higher training test score. In addition, younger people tend to have more experience of manipulating 3D objects in gaming or computer environment than elder ones, r=0.508, N=30, p=0.004.

4.2 Utility of User Perception Measures

To assess the utility of user perception measures, the response of self-efficacy questionnaire (SelfEfficacy) and perceived VE efficacy (PVEefficacy) were included in a multiple linear regression model to predict VE efficacy. Because of potential effects of the independent variable on VE efficacy and the other response measures, LOE3D also was included as a predictor in this model. Finally, memory test score (MTS) was added in the model to account for any susceptibility to cognitive learning outcomes that is essential in quantifying VE efficacy, as shown in equation 2.

$$\mathrm{VEefficacy} = \beta_0 + \beta_1\mathrm{LOE3D} + \beta_2\mathrm{TTS} + \beta_3\mathrm{SelfEfficacy} + \beta_4\mathrm{PVEefficacy} + \beta_{17}\mathrm{MMT} + \varepsilon \tag{2}$$

Results of the standardized regression coefficients analysis indicated that the regression model adequately described self-efficacy believe, F (7.822), p=0.000. The results also shows that at the α=0.05 level, training test score (TTS) (Beta=0.636, p=0.000) is the most important predictor of VE efficacy. Other predictors, memory test score (MMT) (Beta=0.266, p=0.000), perceived VE efficacy (PVEefficacy) (Beta=0.233, p=0.000), and self-efficacy beliefs (SelfEfficacy) contribute to the model slightly (Beta=0.193, p=0.000). However, LOE3D show no contribution to the model (Beta-.000, p=.706).

In addition, results of the Pearson correlation coefficients revealed strong, positive and significant relationships between VE efficacy and TTS (r=0.888, N=30, p=0.000), and between VE efficacy and MMT (r=0.766, N=25, p=0.000). Moreover, the result also shows a moderate, positive and significant linear relationship between VE efficacy and SelfEfficacy (r=0.637, N=30, p=0.000), and between VE efficacy and PVEefficacy (r=0.585, N=30, p=0.000). Interestingly, a moderate, positive and significant linear relationship also found between VE efficacy and LOE3D (r=0.506, N=28, p=0.000). These results shows people who achieve higher on training test (TTS) tend to have higher VE efficacy score. In addition, a moderately weak but positive linear relationship between PVEefficacy and TTS (r=0.384, N=30, p=0.036) suggests that people who perceive VE to be effective achieved higher TTS. A moderate and significant linear relationship also found between LOE3D and TTS (r=0.529, N=28, p=0.529), which suggest that people with high LOE3D tend to perform better than those with low and moderate LOE3D. Interestingly, no significant relationship found between self-efficacy (measured before the training test) and perceived VE efficacy (measured after the training test). Even thought a positive correlation exist between the two, but it is not significant, r=0.177, N=30, p=.175.

4.3 LOE3D on TTS, SelfEfficacy, PVE Efficacy and MMT

One-way analysis of variance (ANOVA) was performed and results show that there was significant effects of LOE3D on task performance, F=7.586, p<.05. Turkey Post Hoc test revealed that subjects performed task better (p<.05) under the moderate LOE3D (mean=82) than under the high (mean=85) and low LOE3D (mean=43). Counter to our expectations, results on SelfEfficacy and PVEefficacy revealed no significant effect (P>.05) of LOE3D. SelfEfficacy and PVEefficacy were observed as dependent measures in this study. People with low LOE3D have similar self-efficacy beliefs and perceive VE to be effective as these with moderate and high LOE3D. Mean score on TTS, SelfEfficacy and PVEefficacy indicate LOE3D have effects on these measures that account for VE efficacy as shown in Figure3. Additionally, one-way ANOVA analysis shows that there is no statistic significant difference on subjects' ability of recall in memory test (MMT) across LOE3D, F=1.852 p>.05. All subjects were able to recall learning task or procedures at high level, regards different ranges of LOE3D (mean>80).

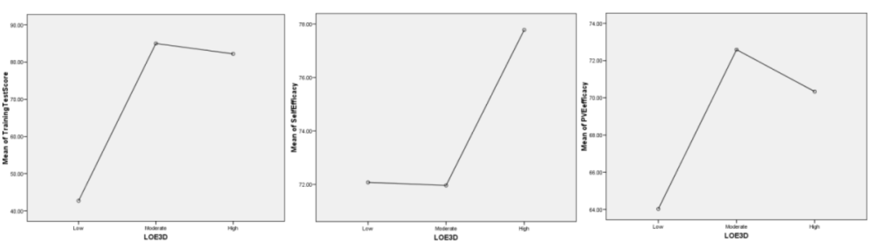

Fig. 3. Effects of LOE3D on TTS, SelfEfficacy and PVEefficacy

4.4 LOE3D on VE Efficacy

ANOVA analysis also shows that there is a statistic significant difference on VE efficacy score across LOE3D, F=4.42, p<.05. Tukey Post Hoc test further suggests that the differences lie between low and high experienced subjects, and no significant difference between moderate and high experienced subjects (p>.05). Mean results shows that moderate and high experienced subjects share similar VE efficacy sore (M>70), and low experienced subjects have lower VE efficacy score (M<60).

Table 1. Effects of LOE3D on VE efficacy

	Low Vs. Moderate	Moderate Vs. High	High Vs. Low
TTS	P<.05	P>.05	P<.05
SelfEfficacy	P>.05	P>.05	P>.05
PVEfficacy	P>.05	P>.05	P>.05
MMT	P>.05	P>.05	P>.05
VE efficacy	P<.05	P>.05	P>.05

4.5 Summary

We found that users with a different range of LOE3D had little to no effect on the self-efficacy and perceived VE efficacy in the VE. As mentioned above, different level of prior experience in manipulating 3D was gathered based on subjective self-report of their expertise. The manipulation of experimental group was affected by such information. With respect to LOE3D on TTS and VE efficacy, LOE3D of high, moderate and low ranges may not have been substantial enough to affect subject's ability to predict their performance and rate VE system efficacy. Besides, subjective perception may not be consistent with objective task performance measures. As in motivational/affect literature, self-efficacy and user attitude (perception) should be used as supplement of objective measures in evaluating VE system performance [10, 15]. Supported by the results of this study, we believe that user perception measures are equally important (if not superior) to assess system efficacy. Even though self-efficacy do not correlated with object measure of task performance significantly well, a positive relationship is detected between PVE and VE efficacy; and both user perception measures of self-efficacy and PVE efficacy are positively and strongly correlated with VE efficacy.

5 Conclusion

Various evaluation methodologies and techniques can be considered and applied for evaluating efficacy of VE systems designed for procedural task training. This paper has discussed issues related to the evaluation of this particular class of applications. Utility of user perception measures of self-efficacy and perceived efficacy based on our proposed evaluation methodology have shown significance in quantifying VE efficacy. The experiment confirms the general hypothesis that a positive correlation exists between subjective and objective measures designed specifically to quantify VE efficacy. Additionally, previous studies have not investigated a model of VE efficacy based on the combined objective measures of task performance and subjective measures of user perception. We also incorporated users previous experience in a computing environment, this past expertise possessed by the test subjects has not been done before when evaluating the effectiveness of a VE. As our research has found more study is required in this direction in order to clearly establish any relationships between self-efficacy, perceived VE efficacy and task performance.

References

1. Bandura, A.: Self-efficacy: the Exercise of Control. W.H. Freeman, New York (1997)
2. Bouffard-Bouchard, T.: Influence of Self-efficacy on Performance in a Cognitive Task. Journal of Social Psychology 130, 153–164 (1990)
3. Davis, F.D.: User Acceptance of Information Technology: System Characteristics, User perceptions and Behavioral impacts. International Journal of Man-Machine Studies 38, 475–487 (1993)
4. DeVellis, R.F.: Scale Development: Theory and Applications, 2nd edn. Sage Publications, Thousand Oaks (2003)
5. Hill, T., Smith, N.D., Mann, M.F.: Role of Efficacy Expectations in Predicting the Decision to Use Advanced Technologies: the Case of Computers. Journal of Applied Psychology 72, 307–313 (1987)
6. Hoffman, B., Schraw, G.: The influence of self-efficacy and working memory capacity on problem-solving efficiency. Journal of Learning and Individual differences 19(1), 91–100 (2009)
7. Isdale, J.: Introduction to VRTechnology. In: IEEE Virtual Reality Conference, p. 302. IEEE Computer Society Press, Los Alamitos (2003)
8. Jia, D., Bhatti, A., Nahanandi, S.: Computer-simulated Environment (VE) for training: Challenge of Efficacy Evaluation. In: The Simulation Technology and Training Conference (SimTecT 2008), p. 63. Simulation Industry Association of Australia (2008)
9. Lin, J.J.W.: Virtual Guiding Avatar: an Effective Procedure to Reduce Simulator Sickness in Virtual Environments. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 719–726. ACM Press, New York (2004)
10. Mantovani, F.: VR Learning: Potential and Challenges for the Use of 3D Environments in Education and Training. In: Giuseppe, R., Carlo, G. (eds.) Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age, pp. 207–226. IOS Press, Amsterdam (2001)
11. Moreno, R.: Does the Modality Principle Hold for Different Media? A Test of the Method-Affects-Learning Hypothesis. Journal of Computer Assisted Learning 22, 149–158 (2006)
12. Nash, E.B., Edwards, G.W., Thompson, J.A., Barfield, W.: A Review of Presence and Performance in Virtual Environments. International Journal of Human-Computer Interaction 12(1), 1–41 (2000)
13. North, M.M., Mathis, J.R., Madajewski, A., Brown, J.T., Cupp, S.M.: Virtual Reality and Transfer of Learning. In: Michael, J.S., Richard, J.K., Gavriel, S., Don, H. (eds.) Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, vol. 1, pp. 634–638. Lawrence Erlbaum Associates, London (2001)

14. Papasratorn, B., Wangpipatwong, T.: The Effects of Self-Efficacy and Attitude on E-learning Outcomes. In: World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education, pp. 226–2270. Association for the Advancement of Computing in Education (AACE) (2006)
15. Riley, J.M., Kaber, D.B., Draper, J.V.: Situation Awareness and Attention Allocation Measures for Quantifying Telepresence Experiences in Teleoperation. Journal of Human Factors and Ergonomics in Manufacturing 14, 51–67 (2004)
16. Salzman, M.C., Dede, C., Loftin, R.B., Chen, J.: A Model for Understanding How Virtual Reality Aids Complex Conceptual Learning. Presence: Teleoperators and Virtual Environments, Special Issue on Education 8(3), 293–316 (1999)
17. Stanney, K.M.: Handbook of Virtual Environments Design, Implementation, and Applications. Lawrence Erlbaum Associates, London (2002)
18. Tesfazgi, S.H.: Survey on Behavioral Observation Methods in Virtual Environments. Research Assignment, Delft University of Technology (2003)
19. Wang, A.Y., Newlin, M.H.: Predictors of Web-student Performance: the Role of Self-efficacy and Reasons or Taking an On-line Class. Journal of Computers in Human Behavior 18, 151–163 (2002)
20. Witmer, B.G., Singer, M.J.: Measuring Presence in Virtual Environments: A Presence Questionnaire. Journal of Presence: Teleoperators and Virtual Environments 7(3), 225–240 (1998)