Examining usability of a virtual reality driving simulator

Maria T. Schultheis, Ph.D.1,2,3
Jose Rebimbas, BA3
Ronald Mourant, Ph.D.4
Scott R. Millis, Ph.D.5

1Drexel University, Department of Psychology
2Drexel University, School of Biomedical Engineering, Science and Health Systems
3Kessler Medical Rehabilitation Research & Educ. Corp.
4Northeastern University, Department of Mechanical Engineering
5Rehabilitation Institute of Michigan

Correspondence and reprint request:
Maria T. Schultheis, Ph.D., Drexel University, Department of Psychology
Email: schultheis@drexel.edu

Preparation of this manuscript and research presented was supported by the National Institute on Disability and Rehabilitation Research grant #H133G000073. The research was conducted at Kessler Medical Rehabilitation Research and Education Corporation.

Abstract:
Statement of Problem: The current study examined specific aspects of usability of a newly developed VR Driver Rehabilitation (VR-DR) system.
Method: Measures of user feedback and user-comfort were examined among 54 participants, 33 individuals with acquired brain injury (20 with traumatic brain injury (TBI) and 13 with cerebral vascular accident (CVA)) and 21 healthy controls (HC). All participants were administered the VR-DR and completed the VR-DR User Feedback Questionnaire.
Results: To examine group differences a one-way analysis of variance (ANOVA) was performed, comparing the User Feedback total score between the three groups. The results indicated that the two clinical populations (TBI and CVA) varied from the non-clinical population (HC). A standard multiple regression analysis revealed that age was the only significant participant factor that contributed to the differences in user feedback ratings. Finally, consistent across the three groups, a distinct relationship was found between the self-reported user rating and the onset of simulation sickness.
Conclusions: The current findings indicated that individuals with TBI and CVA provided less favorable user-feedback ratings than HC in the use of a new VR-DR system. This difference was not accounted by differences in gender, education or cognitive status, and only slightly accounted for by age. Delineating these various aspects of user-feedback can assist in identifying potential confounds in VR-DR performance and help refine the application of the VR-DR for clinical decision-making.

Key words: virtual reality, driving, cognition, user-feedback, usability, human factors
INTRODUCTION

In recent years, the use of virtual reality (VR) technology in medicine has increased in both numbers and applications. The use of this technology has been applied for clinical, research and educational purposes. Within rehabilitation medicine, studies using VR have provided encouraging findings that underscore the unique match between rehabilitation and this technology. Examples of VR rehabilitation applications include environments for the evaluation and retraining of physical abilities (Burdea, Deshpande et. al., 1997; Jack, Boian, et al, 2001), component cognitive abilities (Brooks and Rose, 2003; Lengenfelder, Schultheis et al., 2002; Schultheis, Himelstein et al., 2002) and functional activities of daily living (ADL). The latter includes applications such as the use of a “Virtual Kitchen” for retraining safety skills (Christiansen, Abreu et al., 1998; Zhang, Abreu et al., 2003), the use of a “Virtual City” for evaluating and retraining pedestrian skills (McComas, MacKay et al., 2002), the use of a “Virtual Office” for examining vocationally relevant cognitive abilities (Matheis, Schultheis et al., 2003) and the use of VR Driver simulation for evaluating driving capacity following neurological compromise (Schultheis & Mourant, 2001; Wald & Lui, 2001). The results of these studies have provided evidence of the various assets offered by this technology.

Despite the growing number of VR applications being developed for rehabilitation populations, very little work has been conducted to examine aspects of usability of these new virtual environments. In general, the focus of medical and/or clinical VR work has been in the development of new testing environments and scenarios (Hix and Gabbard, 2002; Hix, 2005; Scerbo, 2005). However, knowledge gained through examining aspects of usability, which describes the quality of a tool that makes it easy to use, can provide critical information for the development of both useful and usable VR clinical applications.

Specifically, in rehabilitation applications of VR, the users, which in most studies have included individuals with cognitive and/or physical disabilities, may have distinctive needs and considerations that are not typically encountered with non-clinical populations. For example, evaluating the cognitive load/demand for using the device, determining the accessibility and adaptability of the device to meet the needs of individuals with different types of physical disabilities and identification of discomforts/side effects that may present unique vulnerabilities to clinical populations. The inclusion of usability methodology could help to address these questions and contribute to establishing the overall effectiveness of the technology/tool (i.e. virtual environments) in generating the intended outcome (e.g. rehabilitating functional everyday activities).

BACKGROUND

By definition usability, a key concept in human-computer interaction (HCI), addresses the relationship between tools and their users and purports that in order for a tool to work, the intended users must be able to effectively use the tool. Usability is the quality of a tool that makes it easy to learn, easy to use, easy to remember, makes it error tolerant and subjectively pleasing (http://www.usabilityfirst.com). In sum, from the user’s perspective, a technology/tool with high usability can make the difference between 1) the user performing a task accurately or not, 2) the user enjoying the process or becoming frustrated and, 3) the user’s acceptance or rejection of the new technology/tool.

In the application of VR technology to rehabilitation medicine, all these factors and the unique challenges that rehabilitation users are faced with (e.g. physical or cognitive deficits) should be considered in the development process. Additionally, in the development of any clinical tool, there are two intended users that require consideration: the patient and the clinician.
As such, evaluating the usability of a new technology (e.g. VR) requires input from both the patient perspective (e.g. comfort, ease of understanding instructions, side effects) and the clinician’s perspective (e.g. ease of use, cost-effectiveness, system complexity).

Addressing usability factors can increase a clinician’s acceptance of the technology/tool and facilitate the overall integration of the technology/tool into clinical application. A high level of usability would indicate that issues related to user interface have been addressed and that the measurement or information provided by the tool is an accurate reflection of the behavior of interest, and thereby usable for clinical recommendations. Additionally, delineating what aspects of performance are a result of the technology/tool, what aspects are the user’s actual responses and what aspects may be a result of the interaction between user and technology/tool could help establish the validity of the measurements provided.

The evaluation of usability can be accomplished through user-feedback, task analysis, evaluation of user interactions and usability inspection. Many of these are typically addressed prior to the development of new technologies/systems and are often dependent on the type of system (e.g. task analysis, errors). Others are dependent on human subject testing (e.g. user-feedback, performance assessment) (Preece, Rogers, Sharp et al, 1994). For the current study, user-feedback was selected as the first aspect of usability to be examined due to the dearth of research that has explored the client/patient’s perspective, the post-VR development nature of the study and because such feedback could provide critical information for further investigation.

The current study examined aspects of usability in a VR Driver Rehabilitation (VR-DR) system designed specifically for research purposes. The VR-DR is a low-cost driving simulation system developed by Schultheis et al (2001) for the assessment of driving capacity following acquired brain injury. Specifically, the VR-DR is a computer based system that uses a head mounted display (HMD) unit to visually present computer-generated driving environments through which users can “drive through” using a commercially available steering wheel and foot pedal (Microsoft Thrust-Master) (see Figure 1). The VR-DR includes a 30-minute virtual driving route that allows the user to drive through a variety of driving environments (e.g. residential zone, highway, commercial zones) in either daytime or nighttime driving conditions. In addition to the visual feedback, the VR-DR provides auditory (e.g. sound of car engine) and tactile feedback (e.g. force feedback from the steering wheel) to increase the VR experience for the user. The VR-DR also allows the presentation of different driving challenges (e.g., pedestrian suddenly crossing the street). A total of six challenges, developed based on clinical experience, can be presented throughout the route. The VR-DR automatically generates four core driving performance measurements; 1. speed of vehicle, 2. lane position of vehicle, 3. degree of head turning and 4. distance from target items (e.g. stop sign, traffic light). These core measurements can be combined to generate more specific driving behavior variables (e.g., scanning behavior at stop sign). The VR-DR was designed to address limitations in current tools for the assessment of driving capacity. Preliminary findings using the VR-DR have demonstrated the concurrent validity of VR-DR measures (Himelstein et al, 2003, Schultheis et al 2004). Specifically,
significant correlations have been reported between VR-DR measures and cognitive tests demonstrated to be relevant to driving capacity. To date, the VR-DR has been administered to individuals with traumatic brain injury (TBI) and cerebral vascular disorder (CVA).

In order to examine aspects of usability related to the VR-DR a user-feedback questionnaire was developed (see Appendix A). The questionnaire was designed to assess participant feedback on 1). comfort with mechanical aspects (e.g. steering wheel, foot pedals, HMD) of the VR-DR, 2). feedback on features of the virtual environment (e.g. lane markings, traffic signs) and 3). subjective comfort with use of the VR-DR. The questionnaire was administered to clinical (i.e., TBI, CVA) and non-clinical (i.e., healthy controls) samples.

In addition, because simulation sickness is a common secondary effect of VR exposure and previous work employing head mounted displays and motion activities (i.e., driving) have demonstrated a high incidence of simulation sickness (Schultheis, 2005), the current study sought to examine the relationship between user-feedback and incidence of simulation sickness as a secondary analysis. One potential utility of the User-Feedback Questionnaire would be its use as a predictor of simulation sickness.

The goals of this study were threefold. First, 1) to identify differences in user feedback between the clinical and non-clinical users, 2) evaluate if any specific characteristics of the clinical group were predictive of the user-feedback and 3) examine the relationship between user-feedback and simulator sickness onset.

METHODS

Participants: A total of 54 participants were included in the study, 33 individuals with acquired brain injury (ABI) and 21 healthy controls (HC). Of the ABI participants, 61% (n=20) experienced a moderate to severe traumatic brain injury (TBI), while the other 39% (n=13) experienced a cerebrovascular accident (CVA) or stroke. Medical records obtained from rehabilitation hospitals and/or treating physicians confirmed diagnosis of injury for ABI participants. ABI subjects were recruited through the Kessler Driving Evaluation Program in both the Saddle Brook and West Orange facilities and via advertisements through the Brain Injury Association of New Jersey. HC subjects were recruited from hospital staff, friends/family of participants, and prior research study databases within the laboratory. Participants with prior, history of severe psychiatric disturbances, extreme motion sickness, substance abuse or history of prior TBI/CVA or any major medical/neurological condition were excluded from the study. Individuals that required the use of assistive driving devices were not included in the study. At the time of testing all participants held a valid driver’s license in the states of New Jersey, New York or Connecticut and were under the age of 68. Reports from the Department of Motor Vehicles were obtained for verification of driving status for all study participants. Individuals with a history of loss of driving privileges and/or reckless driving were not included in the study. All participants were required to have a minimum of one year of continuous driving experience, with ABI participants meeting this requirement prior to injury. Furthermore, all participants were required to meet the minimum visual requirement for their licensing state. Specifically, this was a minimum of 20/50 visual acuity rating in at least one eye for New Jersey and a minimum of 20/40 visual acuity rating with or without corrective lenses in both eyes for New York and Connecticut. Finally, in accordance with legal and clinical requirements all ABI participants who were on anticonvulsant medications were required to be seizure free for at least one full year prior to testing.

The total sample included 37 male and 17 female participants. The average age was 41.7 years and average education was 15.1 years. Eighty-three percent of the sample was Caucasian
(n= 45), 9% Hispanic (n=5), 4% African American (n=2) and 4% identified themselves as other (n= 2). At the time of testing, 87% (n= 47) reported being active drivers, while 13% (n=7) reported they had limited or discontinued driving. For the purposes of this study “active” driving was defined as a minimum of one driving experience per week. For most individuals not actively driving at the time of testing, limited financial resources (e.g. couldn’t afford vehicle, auto insurance) were the primary reason for discontinued driving. For driving experience, defined as total number of years driving, was an average of 23.7 years for the group. Table 1 summarizes demographics for the three groups.

Table 1 Demographic characteristics of sample groups

<table>
<thead>
<tr>
<th></th>
<th>Healthy (n=20)</th>
<th>TBI (n=21)</th>
<th>CVA (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>36.2 (12.5)</td>
<td>40.9 (13.8)</td>
<td>52.7 (11.9)</td>
</tr>
<tr>
<td>Range</td>
<td>[21-64]</td>
<td>[20-61]</td>
<td>[30-68]</td>
</tr>
<tr>
<td>Education (yrs)</td>
<td>16.3 (2.1)</td>
<td>14.5 (2.5)</td>
<td>14.0 (2.4)</td>
</tr>
<tr>
<td>Time Post Injury (months)</td>
<td>- -</td>
<td>77.4 (76)</td>
<td>17.2 (10.8)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Healthy</th>
<th>TBI</th>
<th>CVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>14</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>Females</td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. M= mean, SD= standard deviation, f= frequency.

Defining cognitive status. In order to define cognitive status for participants in the study, a composite cognitive score was calculated using performance scores from three neuropsychological tests that target specific cognitive domains (working memory, attention and executive functioning) that have been demonstrated to be relevant to driving ability among neurologically compromised populations (Rizzo et al, 2004; Schultheis et al, 2001; Sivak et al, 1980; van Zomeren et al., 1987). This included the WAIS-III Digit Symbol, the Paced Auditory Serial Addition Test (PASAT) and the Trail Making tests. A composite cognitive status score was calculated by computing and summing the z-scores of performance on the individual tests.

Procedures: The current study is a component of a larger study which included numerous questionnaires, cognitive testing and administration of the VR Driver Rehabilitation (VR-DR) program. Prior to participation, individuals were screened by phone for inclusion/exclusion criteria. Those meeting criteria were then scheduled for a testing session which lasted approximately 4-5 hours. At the initiation of the testing session, all participants completed an Institutional Review Board consent form and required HIPAA authorization forms.

All participants were provided with a practice trial on the VR-DR. The purpose of this trial was to provide participants with an introduction to the VR-DR, allow them to familiarize themselves with the various components of the VR-DR [i.e., steering wheel, foot pedals, head-mounted display (HMD)] and to allow them the opportunity to experience “driving” using the VR-DR.
Specifically, the practice consisted of two trials, during which time participants were instructed to “drive thru” a simple virtual road using the VR-DR. Prior to driving initiation, participants were instructed on how to use the steering wheel and foot pedals and were provided specific driving directions (i.e., “stay in the right hand lane and maintain a speed of 35 mph”). Each practice trial required participants to come to a full stop on three separate occasions, two on curvy portions of the route and one on a straight portion of the route. The first practice trial was completed using a flat-screen (no head mounted display) version of the simple virtual road, while the second practice trial was administered with participants wearing the HMD and head tracking device.

Following this, all participants completed the VR-DR User Feedback Questionnaire. This questionnaire consists of a total of eight statements about aspects of the VR-DR. For each item participants were required to assign a numeric value to each statement ranging from one to nine, with one indicating strong disagreement with the statement and nine indicating strong agreement with the statement. The scale generates an overall rating score (User Feedback Total Score), with a range of 8 – 72, where higher scores indicate favorable VR-DR use.

Following this, all participants were administered the remainder of the protocol, which include completion of cognitive tests, completion of the VR-DR Test Trials (three 30-minute virtual routes) and administration of simulation sickness monitoring questionnaires. Fatigue and other aspects of potential cognitive overload were monitored via self-report during the testing session. All subjects were provided both structured and requested rest periods during the testing.

**RESULTS**

![Graph showing Group Differences](image)

**Group Differences:** To quantitatively examine group differences in user-feedback rating of the VR-DR, a one-way analysis of variance (ANOVA) was performed, comparing the User Feedback total score between the three groups [TBI, CVA and HC]. Results revealed a main effect for Group \([F(2,54)= 5.02; p=.010]\). Post-hoc Tukey analysis revealed that the HC group reported significantly more favorable VR-DR ratings (M= 55.1) than both the TBI group (M= 46.6) and the CVA group (M=43.5). No significant differences were seen between the TBI and CVA group (see Figure 2).

A qualitative analysis of the individual User-Feedback Questionnaire items was conducted to examine the specific aspects of the VR-DR that were viewed unfavorably. In examining items related to the comfort of the mechanical aspects of the VR-DR, the HC and TBI group averaged favorable ratings (score greater than 5) for the use of the steering wheel and foot pedals, while
the CVA group averaged a “neutral” (score of 5) rating. Use of the HMD received somewhat less favorable ratings by all three groups. None of the groups reported difficulties with the noise of the virtual environment or difficulties in seeing lanes marking within the virtual driving environments.

Most group differences were observed in items which include interpretive statements. Specifically, the CVA group averaged the highest reports of “feeling that the vehicle was going off the road” and both the TBI and CVA group averaged low ratings for the statement “the VR-DR feels like driving a car”. The HC group averaged high ratings for feeling overall “in control of the vehicle” followed by the CVA and the TBI group.

User Feedback variance explained by participant demographics: A standard multiple regression analysis was performed between the User-Feedback total score as the dependent variable and four predictor variables: age (in years), education (in years), gender, and a cognitive composite score (described previously). Overall, the model accounted for a modest amount of the variance in User-Feedback scores (8.1% adjusted). Further analysis of the predictor variables indicated that only one variables (age) contributed significantly to the prediction of the User-Feedback score (Table 2).

### Table 2: Summary for participant variables predicting total score on User-Feedback Questionnaire

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SEB</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>-0.25</td>
<td>0.12</td>
<td>-0.29</td>
<td>0.048*</td>
</tr>
<tr>
<td>Gender</td>
<td>-2.0</td>
<td>3.7</td>
<td>-0.08</td>
<td>0.587</td>
</tr>
<tr>
<td>Education</td>
<td>0.90</td>
<td>0.75</td>
<td>0.18</td>
<td>0.233</td>
</tr>
<tr>
<td>Cognitive Status</td>
<td>1.20</td>
<td>1.5</td>
<td>0.12</td>
<td>0.444</td>
</tr>
</tbody>
</table>

Note. R2=.16 (n=54, p=.09)
* p < .05

Utility of the User Feedback total score:
A direct logistic regression analysis was performed on the development of simulation sickness as outcome (yes/no) and the User-Feedback total score as the predictor variables. A test of the full model with the one predictor against a constant-only model was statistically reliable, $\chi^2 (1), N = 54) = 8.71, p = .003$, indicating that the User-Feedback total score reliably distinguished between patients who experienced simulation sickness and those who did not. This model appeared well-calibrated (i.e., probabilities reflecting the true outcome experience in the data), Hosmer-Lemeshow $\chi^2 (8) = 6.56, p = .59$. Table 3 shows the regression coefficients, Wald statistics, odds ratios, and 95% confidence intervals for the odds ratio. For every increase of 10 points on the User-Feedback total score, the risk of developing simulation sickness decreases by 56%.
### Table 3: Summary of logistic regression analysis predicting onset of simulation sickness

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Odds Ratio</th>
<th>Wald statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-Feedback Total Score</td>
<td>-.081</td>
<td>.029</td>
<td>.923</td>
<td>7.5**</td>
</tr>
</tbody>
</table>

**p < .01

**DISCUSSION**

The current study reflects initial work examining specific aspects of the usability of a newly developed VR driving simulation system. Measures of user feedback and user-comfort were examined among three groups of participants, healthy controls, persons with traumatic brain injury and persons with cerebral vascular disorder. Findings indicated differences in the overall user rating of the VR-DR, suggesting that the two clinical populations (TBI and CVA) varied from the non-clinical population (HC) in their overall comfort with using the VR-DR. Further evaluation of the data, indicated that the greatest group differences were seen in the subjective aspects of user-comfort, such as how realistic the VR-DR felt. The group differences were less pronounced in regards to rating the mechanical aspects of the system (such as using the steering wheel and foot pedals), although the CVA group reported marked difficulties with the use of the head mounted display.

Analysis to examine what specific participant factors may have contributed to the differences in user feedback ratings revealed that only age was a significant predictor. Other important demographic factors (i.e. gender, education, cognitive status) did not predict user-feedback, underscoring the complexity of this dimension. While these findings are consistent with previous studies that have identified age as an important issue in acceptance and usage of new technologies (Morris and Venkatesh, 2000), future research including measurement of participant related variables, such as suggestibility and personality, would be beneficial. Interestingly, cognitive status did not play a role in determining user-feedback rating for this sample. This finding is encouraging and gives support for the continued use of the VR-DR with these clinical populations.

Finally, not surprising and consistent across the three groups, a distinct relationship was found between the self-reported user rating and the onset of simulation sickness. Specifically, as users discomfort increased, the likelihood of simulation sickness also increased. These findings support two important considerations. First, it may be possible that the VR-DR User Feedback Questionnaire may serve as a screening tool and help establish a cut-off for determining those individuals who may or may not be appropriate for VR exposure. Such a tool would be beneficial for reducing the incidence of simulation sickness and improve our ability to appropriately match technology and user. For example, it may be that individuals with stroke may be at higher risk for simulation sickness, the use of a screening tool could help determine whether retraining in a virtual environment may be appropriate for a particular individual. Second, the identified relationship between the User-Feedback Questionnaire and the onset of simulation sickness, suggests the need to further examination into what aspects of user-comfort may be of relevance to simulation sickness. Future work, examining incremental adjustments to user-comfort and its impact on simulation sickness onset, could help establish what aspect of VR-exposure is most effective in reducing simulation sickness.
One of the limitations of the current study is that explicit usability measurements were not identified apriori and subsequently, these initial analyses provide global and not specific information regarding the usability of the VR-DR. For example, additional performance measurements (e.g. errors in using hardware, time to learn system use) could have been defined and quantified to better establish the impact of these features. Future work, incorporating the use of established user-feedback measurements and targeting specific aspects of the VR-DR, such as the donning and use of the HMD, would help further clarify usability issues. Additionally, it would be important for future studies to expand these findings by evaluating additional factors (e.g. level of computer knowledge/exposure) that may influence ratings of VR use.

In conclusion, as VR rehabilitation applications continue to evolve from research project into a platform for medical applications, as this continues to progress the significance of evaluating usability should not be overlooked. Obtaining user feedback in a systematic manner, defining the user-interface and involving users in the development process are all strategies well-defined in usability engineering that can assist rehabilitation specialists in developing and integrating new VR tools.

As the use of VR tools becomes more common in rehabilitation settings and reliance on VR measures for clinical decisions or recommendations increases, it becomes incumbent upon researchers and clinicians to evaluate not only traditional properties of reliability and validity, but also the usability of these new tools. As such future work should include continued evaluation of VR-DR measures, including validation against other measures of driving capacity, examining the repeatability of collected measures and designing studies targeting more specific aspects of usability. Ensuring that the measures we are collecting from the VR are not confounded by the use of the technology can increase the acceptance and integration of this technology into clinical practice.

REFERENCES


APPENDIX A
VIRTUAL REALITY DRIVING SYSTEM - USER FEEDBACK

Subject ID: ______________ Date:__________________ Examiner __________________________

DIRECTIONS: People differ in their ability to drive comfortably using the virtual reality driving system. This is a scale to measure how you feel about your ability to drive using the virtual reality system following the practice session. Read each statement and provide a number from this scale of 1 – 9 that indicates how true the statement is for you. Do not spend too much time on any statement. Answer honestly.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Disagree</td>
<td>Somewhat Disagree</td>
<td>Neutral</td>
<td>Somewhat Agree</td>
<td>Strongly Agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RATING

1. I feel comfortable using the Virtual Reality Driving System (VRDS) steering wheel.  
2. I can control the speed of the vehicle using the VRDS gas pedal and brake.  
3. I have difficulty seeing the lane markings in the Virtual Reality Driving System.  
4. Driving in the Virtual Reality Driving System feels almost like driving in a car.  
5. The background noise from the Virtual Reality Driving System bothers me.  
6. I think that the car will go off the road often when I drive the VR Driving System.  
7. I feel like I am in control of the car when driving with the VR Driving System.  
8. Wearing the Head Mount Display makes it difficult for me to drive in the VRDS.  