

Running head: Virtual Learning and Memory

IS LEARNING AND MEMORY DIFFERENT IN A VIRTUAL ENVIRONMENT?

Robert J. Matheis, Ph.D.¹

Maria T. Schultheis, Ph.D.^{2,4}

Lana A. Tiersky, Ph.D.¹

John DeLuca, Ph.D., ABPP^{3,4}

Scott R. Millis, Ph.D., ABPP³

Albert Rizzo, Ph.D.⁵

¹Fairleigh Dickinson University, Teaneck, NJ, School of Psychology

²Kessler Medical Rehabilitation Research and Education Corporation, West Orange, NJ,

Rehabilitation Engineering and Analysis Laboratory (REAL)

³Kessler Medical Rehabilitation Research and Education Corporation, West Orange, NJ,

Neuroscience Research

⁴University of Medicine and Dentistry of New Jersey, Newark, NJ, Department of

Physical Medicine and Rehabilitation

⁵University of Southern California, Los Angeles, CA, Integrated Media Systems Center

Word Count: 4909

Corresponding Author:

Maria T. Schultheis, Ph.D.

Kessler Medical Rehabilitation Research
and Education Corporation

1199 Pleasant Valley Way

West Orange, NJ 07052

mschultheis@kmrrec.org

Office: 973-324-3528

Fax: 973-243-6984

Abstract

It has been suggested that virtual reality may provide a medium for producing neuropsychological measures with greater ecological validity. The present study examined the usefulness of virtual reality (VR) to assess learning and memory in individuals with traumatic brain injury (TBI). Twenty TBI participants were compared with 20 healthy controls on their ability to learn and recall 16 target items presented within a VR-based generic office environment. The results indicated that VR memory testing accurately distinguished memory impaired TBI participants from others. Additionally, non-memory impaired TBI participants acquired targets at the same rate as HC participants. Finally, there was a significant relationship between the VR Office and a standard neuropsychological measure of memory, suggesting the construct validity of the task. These findings suggest that the VR Office provides a viable medium for measuring learning and memory. Additionally, the VR Office may have enhanced ecological validity, which ultimately can improve assessment of real-world functioning following TBI.

Is Learning and Memory Different in a Virtual Environment?

Traumatic brain injury (TBI) occurs when mechanical force causes damage to brain tissue resulting in the physiological disruption of brain functioning. Cognitive impairment, a common feature post-TBI, is typically measured with paper-and-pencil based neuropsychological measures. Results obtained on formal neuropsychological evaluation are then interpreted by clinicians to provide appropriate treatment recommendations and to assess how identified neuropsychological strengths and weaknesses will manifest within real-world, functional scenarios such as school or work (Lezak, 1995). This assessment of functional outcome is often of fundamental importance to both ongoing treatment planning and to the accurate prediction of how cognitive impairment will impact long-term functional status (Newman, Heaton, & Lehman, 1978). Thus, the ecological validity or generalizability of neuropsychological test results, beyond the clinical environment is a crucial component of the evaluative process (Kibby, Schmitter-Edgecombe, & Long, 1998).

While numerous studies support the predictive value of standard neuropsychological measures (Newman, Heaton, Lehman, 1978; Sureyya, Dikmen, Machamer, Powell, Temkin, 2003), the capability of these instruments to accurately predict real-world functioning may be diminished by poor ecological validity (Bennett, 2001; Bowman, 1996; Farias, Harrell, Neumann, & Houtz, 2003; Gioia & Isquith, 2004; Ptok, Buller, Kuske, & Hecker, 2004; Sbordone & Long, 1996; Vriezen & Pigott, 2002). Various reasons accounting for why standard measures of neuropsychological functioning may often show poor ecological validity have been discussed (Schultheis, Himelstein, & Rizzo, 2002). One reason is that traditional clinical measures have poor

face validity and *appear* different from the real-world scenarios to which an individual actually functions. This difference in appearance is principally related to the fact that standard tests do not reproduce the rich array of stimuli that an individual experiences in the real-world. The incongruity in appearance between test and environment can impact examinee performance (e.g., motivation and effort). Additionally, standard instruments typically examine isolated components of neuropsychological ability, which is different from real-world scenarios in which individuals often employ a combination of cognitive skills and abilities simultaneously. Thus, on the basis of standard instruments it can be difficult to assess empirically how a measured deficit will manifest outside the artificial, clinical testing environment. For these and other reasons discussed by Schultheis et al., it has been suggested that clinicians re-examine and update their repertoire of neuropsychological tools, with an aim towards achieving greater ecological validity.

One way in which neuropsychological measures might achieve enhanced ecological validity is through the use of virtual reality. Virtual reality (VR) is a technology that enables humans to become immersed in and interact with three-dimensional computer-generated environments by employing integrated computer components (e.g., head mounted display) to deliver a multi-sensory experience that resembles the real-world. Different from two-dimensional, traditional computer programs and simulations, VR allows individuals to attain a greater “sense of presence” as they become immersed within the computer-created environment. Thus, analogous to an aircraft simulator, virtual environments have the potential to provide interactive scenarios that can be useful in both assessing and re-training neurocognitive abilities (Schultheis & Rizzo, 2001).

Within the domain of neuropsychological assessment, a growing literature has begun to investigate the extent to which VR environments might be employed to create instruments with greater ecological validity and possibly yield data that have greater generalizability to real-world scenarios. Specifically, neuropsychological tasks have been administered to individuals while they are immersed within a virtual environment in order to determine how skills and abilities manifest themselves in an environment that appears more similar to the real-world. The opportunity to embed or integrate our standard instruments within a computer-generated environment affords a number of valuable benefits. Specifically, VR-based neuropsychological tasks allow for enhanced task control and response measurement, where the technology allows task stimuli and parameters (e.g., number, order, and speed) to be consistently manipulated and patient responses and behaviors to be closely monitored and automatically recorded.

Additionally, VR-based measures allow for the simultaneous measurement of multiple cognitive abilities, which allow clinicians to measure *complex* sets of skills and behaviors that may relate closely to real-world, functional abilities. This is different from standard instruments, in which *components* or isolated domains of cognitive function often are measured and clinicians combine data to predict real-world performance. Notably, the advantages afforded by VR-based assessments are realized while still maintaining the reliability and empiricism that is associated with traditional formalized assessment.

Additionally, patient safety is maintained as there is minimal danger associated with testing in a virtual environment where the limits of an individual's performance can be assessed without the potential for injury that might occur in a real-world scenario.

A number of studies have examined how cognitive functioning might be assessed by integrating neuropsychological tests into virtual environments (Brooks, Rose, Potter, Jayawarden, & Morling, 2004; Buckwalter & Rizzo, 1997; Elkind, Rubin, Rosenthal, Skoff, & Prather, 2001; Ku et al., 2003; Riva, 1997; Rizzo & Buckwalter, 1997; Zhang et al., 2001 & 2003). These studies have focused on both traditional component cognitive assessment of individual neurocognitive abilities (e.g., memory, executive function, attention) and on more functional complex behavioral assessment in which multiple skills and abilities are measured simultaneously. Overall, the results of these studies have found strong concordance between real-world performance and test performance measured within a virtual environment. Additionally, VR-based assessment measures have been found to reliably distinguish between individuals with neurocognitive impairment and healthy individuals, with accuracy similar to standard measures.

In a recent study of complex behavioral assessment using VR, Zhang et al. (2003) compared TBI patient performance on a standard occupational therapy kitchen evaluation with performance on a VR-based kitchen evaluation and found significant concordance between the two testing environments. Specifically, correlations between tasks performed in the VR kitchen and the real kitchen ranged from 0.30 to 0.63, suggesting the construct validity of the VR task with a standard measure of functional ability. In an earlier study using the VR kitchen, Zhang et al. (2001) found that task performance varied significantly between TBI participants and healthy controls, suggesting the VR-based measure could distinguish between clinical and non-clinical populations.

In a study examining VR for component cognitive assessment of executive functioning, Elkind, Rubin, Rosenthal, Skoff, and Prather (2001) compared the

Wisconsin Card Sort Test (WCST) with the Look For A Match (LFAM) test, a similar VR-based measure of executive functioning. In the LFAM, participants are immersed in a beach scene where they are required to deliver various beach items (e.g., Frisbee) to bathers by eliciting a match rule from examiner feedback. Similar to the WCST, participants match bather preferences for color, number, and type of objects to be delivered and then must switch set in accord with examiner feedback. The authors found that, while the LFAM appeared to be a more difficult task, it tapped the same neurocognitive skills as the WCST. However, the LFAM had the distinct advantage of being able to provide a consistent, controlled, and face valid test environment that may ultimately be more predictive of a real-world functional task. This ability to measure a component of cognitive functioning within a controlled, but more elaborate and visually stimulating testing environment is presently not available with our standard repertoire of neuropsychological instruments.

While there is a growing literature examining the usefulness of VR as a medium for creating assessment instruments, very little work has been done with in the area of VR-based memory tests. Considering the high prevalence of memory dysfunction following neurologic damage (Crosson, Novack, Trenner, & Craig, 1988; Vakil, Golan, Grunbaum, Grosswater, & Aberbuch, 1996), further investigation into VR-based memory measures is warranted. One exploratory study of prospective memory ability in individuals post-stroke found that a memory task integrated into a VR-based virtual bungalow environment was a valid and psychometrically sensitive means for assessing memory ability (Brooks, Rose, Potter, Jayawardena, & Morling, 2004). As with many studies to date, however, the virtual bungalow environment was only two-dimensional,

which can effect the examinee's experience of immersion. Additionally, in Brooks et al. the experimenter navigated and controlled the virtual environment for the participant, which may have further reduced immersion into the virtual environment. While these findings are encouraging, the present study will examine an environment with increased experience of immersion and greater environmental control.

In order to examine the utility of using a VR environment to test memory in individuals with TBI, the present study compared TBI participants with healthy controls on a VR-based memory paradigm. Specifically, participants were immersed within the VR Office, a 3-D virtual environment programmed to resemble a common work office and were administered an open-trial, list-learning test of memory. The primary aim was to investigate how memory performance in a VR environment may differ from standard testing. A second aim was to test the hypothesis that, consistent with established measures, the VR-based memory task was expected to distinguish TBI from HC participants. The final exploratory aim was to provide preliminary data to support the construct validity of the task, by correlating the VR Office memory task with scores obtained on a standard neuropsychological battery.

METHOD

Participants

Participants included 20 individuals with moderate to severe TBI and 20 healthy controls (HC) matched on age, gender, and education. Participants were recruited from the outpatient department of an acute rehabilitation hospital and from a day rehabilitation program, both located in northern New Jersey. Prior to participation, all participants completed a consent form approved by the hospital Institutional Review Board. The mean interval between date of injury and time of testing was 7.3 years ($SD = 5.06$; range, 1-22). At the time of testing, 3 (15%) of participants were receiving post-acute rehabilitation (e.g., physical, occupational, neuropsychotherapy), 12 (60%) were employed, and the remainder were on post-TBI disability. Mean loss of consciousness (LOC) for the TBI group was 16 days ($SD = 17.3$; range, 1-60) and mean Glasgow coma scale (GCS) score was 3.5 ($SD = 0.71$; range, 3-4). TBI participants were injured primarily as a result of motor vehicle accidents (60%), falls (30%), and other traumatic injuries (10%). HC participants were recruited from hospital staff and from local university students responding to recruitment postings. Participants with prior neurologic, psychiatric, or substance abuse histories or a documented learning disability were excluded. There were no statistically significant differences in age ($F(1, 38) = 3.40$, $p = 0.07$), gender ($X^2(1) = 0.40$, $p = 0.53$), or education ($F(1, 38) = 3.03$, $p = 0.19$) between the 2 groups.

Apparatus

The VR Office, a 3-D virtual common office environment, runs on a Dell Inspiron 8100 laptop with a Pentium III, 1.13GHz processor, a NVIDIA GeForce 2 video card and

a 15-inch monitor. Three-dimensional visual imagery is presented using 5TH Dimension Technologies (5DT) 800 Series Head Mount Display (HMD) unit. The HMD is attached to The Flock of Birds (FOB) position and orientation measurement system (Ascension Technology Corporation), which transmits head rotation information to the laptop in order to provide real-time updating of the virtual environment. The FOB allows for six degrees-of-freedom and transmits between 20 and 144 measurements per second.

The VR Office software is a 3-D virtual environment programmed to simulate a generic office space. As depicted in Figure 1, the VR Office is programmed to contain common office items such as a desk, a computer, file cabinets, and a phone. During immersion in the VR Office, participants are seated at a desk and have a complete 360-degree view of the office environment.

Insert Figure 1 about here

Measures

Outcome measures consisted of the VR-based learning and memory task, a complete battery of standard neuropsychological measures, and the modified Simulator Sickness Questionnaire (m-SSQ).

VR Learning and Memory Task

The VR Office was programmed to contain 16 target items. The 16 targets consisted of 8 common office items (e.g., phone) and 8 uncommon office items (e.g., blender) that typically are not found in an office setting. A combination of common and uncommon items were included to prevent score inflation by simply naming ordinary

office items during the learning and memory task. Common and uncommon items were counter-balanced in terms of placement throughout the VR Office, so that an equal number were located on either side of the participant.

The VR Office task began when the participant placed the HMD on their head and was immersed within the office environment. In order to standardize the experience, the examiner immediately provided a scripted one-minute orientation to the environment by identifying the four visible quadrants of the office and asking the participant to rotate their head to see them and become comfortable within the environment.

The actual learning and memory task in the VR Office was a verbal list-learning paradigm, which required the participant to learn the 16 target items, depicted among numerous other office distracters (e.g., computer, file cabinet), during a sequence of sequential exposures to the VR Office environment. Specifically, participants were immersed within the office and provided with a maximum of 12 acquisition trials to learn all 16 targets. During each exposure, the participant was asked by the examiner to locate each of the target items, presented in a random sequence to avoid serial effect. Actual VR exposure time varied for each participant, however, participants were provided with only enough time to locate items. Participants were not permitted to continue to explore the environment beyond the time needed to locate each target. Between each exposure, participants removed the HMD and were asked to recall only items listed by the examiner (i.e., target items). Each exposure and subsequent recall constituted one learning trial. Participants continued with VR Office exposures until either all 16 target items were recalled across two consecutive learning trials (i.e., the learning criterion) or until the maximum of 12 learning trials were reached.

Initial acquisition was measured as the number of trials required to meet the learning criterion (i.e., trials to criterion). Short-term recall was measured by assessing the number of items that could be recalled spontaneously following a 30-minute delay. As a measure of item recognition, after 30-minute recall, participants were re-immersed in the VR Office, but during this exposure all target items were removed. Thus, participants were able to see only the core office depicting only the generic office and distracter items. The examiner then provided participants with an electronic pointer to “point” to areas within the VR Office. As the examiner listed target items participants were asked to point to the specific location where each target item previously was depicted. The total number of correctly identified item placements yielded a measure of recognition.

Long-term recall of target items, was assessed by contacting available participants approximately 24 hours following completion of the study and asking them to recall spontaneously items learned during the experiment. As a measure of recognition during this phase, participants were then read a list containing the target and two related distracter items and were asked to identify the correct target. These procedures provided measures of long-term recall and recognition, respectively.

Neuropsychological Measures

The following neuropsychological measures were employed to assess domains of neurocognitive function: The Wechsler Abbreviated Scale of Intelligence (WASI; Harcourt Brace Jovanovich, 1999) was administered as a measure of general psychometric intelligence. The digit span and digit symbol-coding subtests from the Wechsler Adult Intelligence Scale 3rd edition (WAIS; Psychological Corporation, 1997)

assessed simple auditory attention and visual search and scanning abilities, respectively. Visuomotor tracking and processing speed were measured with the Trail Making Test (parts A and B; Reitan, 1955). The Wisconsin Card Sort Test (WCST; Heaton, 1981) was used as a measure of executive functioning. Visual confrontation naming and visual perceptual organization were measured with The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) and the Hooper Visual Organization Test (HVOT; Hooper, 1983), respectively. The Controlled Oral Word Association Test (COWAT; Spreen & Strauss, 1998) assessed verbal fluency. The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) assessed verbal learning and memory and the Brief Visuospatial Memory Test-Revised (BVM-T-R; Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996) measured visuospatial memory. Administration and scoring of all tests were in accordance with standard published procedures

Other Measures

Demographic and injury-related characteristics were collected with self-report questionnaires. The modified Simulator Sickness Questionnaire (m-SSQ) was used to predict likelihood of sickness prior to VR exposure and presence of symptoms following exposure. The m-SSQ contains two parts. The first contains Likert-style questions that predict the likelihood of simulator sickness by asking about past experiences with motion sickness (e.g., flying). Participants with potential to develop simulator sickness were informed and given the option to end participation. The second component of the measure presents 29 symptoms of simulator-sickness on which the participant rates severity both pre and post-exposure to determine if symptoms have emerged during the study.

Procedure

Each testing session lasted approximately two to three hours. Participants first completed self-report questionnaires and the m-SSQ. Next, participants were administered either the VR Office task or the neuropsychological test battery, which were counter-balanced in sequence to control for fatigue. In order to avoid interference, tests of verbal memory (i.e., VR Office, CVLT) were spaced apart maximally within the battery. Immediately following the VR Office task, the m-SSQ again was administered to assess simulator sickness. None of the participants reported simulator-related symptoms. Prior to leaving the laboratory, participants were debriefed regarding the purpose and goals of the study.

RESULTS

Among the 20 participants with TBI, 20% ($n = 4$) of individuals were unable to meet the VR Office task learning criterion, whereas 100% of the HC participants met the criterion. Thus, three groups were identified: individuals with TBI who met the criterion (TBI-MET, $n = 16$), individuals with TBI who did not meet the criterion (TBI-NOT MET, $n = 4$), and HC participants ($n = 20$). Demographic characteristics of these groups are presented in Table 1.

 Insert Table 1 about here

-----**VR Office Learning Measures: Initial Target**

Acquisition

To examine group differences in initial acquisition of target stimuli, mean number of trials to criterion were compared between the TBI-MET and HC groups. TBI-NOT MET participants were excluded as, by definition, they did not meet the learning criterion. Results indicate that the TBI-MET and HC groups were nearly identical in the number of trials required to learn the 16-items presented, $F(1, 34) = 0.00, p = 0.98$. As depicted in Figure #2, the TBI-MET group required an average of 3.94 trials ($SD = 1.61$; range, 2-8), whereas the HC group required an average of 3.95 trials ($SD = 1.91$; range, 2-10) to meet criterion.

 Insert Figure 2 about here

VR Office Memory Measures: Target Recall

VR Office recall performance among the three groups was compared with a 3(Group) X 2(Delay) repeated measures analysis of variance (ANOVA). Results are shown in Figure #3. There was a significant main effect for Group ($F(1, 29) = 43.2, p < 0.001$), and post hoc Tukey tests revealed that the TBI-NOT MET group recalled significantly fewer items ($M = 8.25, SD = 4.57$) than both the TBI-MET group ($M = 14.31, SD = 1.70$) and the HC group ($M = 15.35, SD = 1.04$). The TBI-MET and HC groups did not differ significantly in overall item recall. There also was a significant main effect for Delay ($F(1, 29) = 43.2, p < 0.001$), where as expected when data was collapsed across the 3 groups, significantly fewer items were recalled at the 24-hour delay ($M = 7.85, SD = 0.72$) than at the 30-minute delay ($M = 12.67, SD = 0.41$).

 Insert Figure 3 about here

Notably, the interaction of Delay and Group also was statistically significant ($F(2, 29) = 4.23, p < 0.05$), where performance of the TBI-NOT MET group varied over time. Interestingly, while the TBI-NOT MET group recalled significantly fewer items than the other 2 groups at the 30-minute delay, this difference disappeared at 24-hour delay, when the groups did not differ significantly.

VR Office Memory Measures: Target Recognition

The 3 groups were compared on VR Office recognition performance with a 3(Group) X 2(Delay) repeated measures ANOVA. A significant effect for Group again was found for the total number of items correctly recognized, ($F(1, 29) = 48.4, p < 0.001$)

and post-hoc Tukey tests indicated that the TBI-NOT MET group ($M = 8.25$, $SD = 1.0$) recognized significantly fewer items than both the TBI-MET ($M = 12.5$, $SD = 0.55$) and HC groups ($M = 12.6$, $SD = 0.50$). The TBI-MET and HC groups did not differ significantly in number of items recognized.

As expected, there was a significant main effect for Delay, $F(1, 29) = 48.4$, $p < 0.001$), where across groups a significantly greater number of items were recalled following the 30-minute delay ($M = 13.5$, $SD = 0.30$) than following the 24-hour delay ($M = 8.80$, $SD = 0.70$). This observed change in recognition performance over time did not differ significantly among the groups, as indicated by the lack of a significant Delay X Group interaction.

Comparison of Performance on Neuropsychological Measures

In order to discern the pattern of neurocognitive strengths and weaknesses and the impairment level of our TBI groups relative to HCs, differences in performance on neuropsychological measures among the 3 groups were analyzed with one-way ANOVA (shown in Table 2). The results indicated substantial group differences in performance on neurocognitive measures, with group equivalency observed only on the block design and matrix reasoning subtests of the WASI. These two subtests contribute to the WASI estimate of Performance IQ, in which there also was no significant between-group differences.

Insert Table 2 about here

Overall, the HC group significantly ($p < 0.01$) outperformed the TBI-NOT MET and TBI-MET groups on measures of overall psychometric intelligence (WASI, FSIQ), expressive vocabulary (WASI, vocabulary), visual search and scanning (WAIS-III, digit symbol-coding), verbal learning and memory (CVLT: trial 5, short delay-free recall, long delay-free recall), visuospatial memory (BVRT: total recall, delayed recall), psychomotor speed and complex motor sequencing (TMT, Part B, time), verbal fluency (COWAT). Further, the HC group significantly ($p < 0.01$) outperformed the TBI-NOT MET group on measures of simple auditory attention (WAIS-III, digit span), psychomotor speed and simple motor sequencing (TMT, Part A, time), executive functioning (WCST, categories completed), confrontation naming (BNT), and visual perceptual organization (HCVOT). The TBI-MET group significantly ($p < 0.01$) outperformed the TBI-NOT MET group on measures of verbal learning and memory (CVLT: trial 5, short delay-free recall, long delay-free recall), visuospatial memory (BVRT: total recall, delayed recall), psychomotor speed and simple motor sequencing (TMT, Part A, time), and confrontation naming (BNT). Taken together these results suggest significant neurocognitive impairment for both TBI groups relative to the HC groups and isolated areas of severe impairment for the TBI-NOT MET group relative to the TBI-MET group.

Comparison of TBI Groups

To examine for potential factors between the TBI-MET and TBI-NOT MET groups that may have impacted study performance, the two groups were compared on demographic injury characteristics. Results indicated that the two groups did not differ significantly in days LOC (TBI-MET: $M = 12.66$, $SD = 15.3$; TBI-NOT MET: $M = 29.5$,

$SD = 20.5$) or in years post-injury (TBI-MET: $M = 7.5$, $SD = 5.4$; TBI-NOT MET: $M = 6.5$, $SD = 4.2$).

Construct Validity

As an exploratory analysis and to provide preliminary data to support the construct validity of the VR Office task as a measure of learning and memory, recall indices from the VR Office and the CVLT were correlated. Specifically, Pearson bivariate correlation analyses were used to compare the CVLT Total Recall (i.e., List A, sum of Trials 1 - 5) and Long Delay-Free Recall indices with the total number of words learned at Trial 5 (equivalent to Total Recall, CVLT) and 30-minute delay recall from the VR Office task. The results indicated that while CVLT Total Recall did not correlate significantly with Trial 5 Recall from the VR Office, there was a significant positive relationship between CVLT Long Delay-Free Recall and VR Office 30-minute Recall, $r(39) = 0.70$, $p < 0.001$), with 49% variance shared between the two indices. This result provides preliminary evidence supporting the convergent validity of the VR Office with an established measure of learning and memory.

DISCUSSION

The purpose of the present study was to investigate the utility of using VR to assess learning and memory in people with traumatic brain injury in order to ultimately craft instruments that have improved ecological validity. The approach utilized was to integrate a traditional list-learning paradigm, similar to the CVLT, within the VR Office and to compare learning and memory abilities among people with TBI and healthy controls. Additionally, exploratory analyses were completed to find preliminary evidence in support of the construct validity of the VR Office task as a measure of learning and

memory. Prior to conducting principal analyses, the three groups performance's on standard neuropsychological measures were examined and it was determined that the TBI groups were significantly impaired relative to the HC group.

The results showed that the VR Office provides an effective medium for assessing post-TBI memory ability. Specifically, significant learning differences between TBI and HC participants were found, in that 20% of TBI participants (TBI-NOT MET) were unable to learn successfully all targets compared to 0% of the healthy group. This finding is consistent with DeLuca, Schultheis, Madigan, Christodoulou, and Averill (2000) who also employed an open-trial list-learning task and found that a proportion of only their TBI participants were unable to learn all of the items presented. Notably, however, in the present study, when actual rate of learning between TBI that met criteria and HC participants were compared, the number of trials required to learn items was nearly identical between the two groups. Thus, the expected acquisition impairment that often accompanies moderate to severe TBI did not manifest during testing with the VR Office.

While this finding is atypical within the TBI learning and memory literature, the present study employed an instrument that presented target stimuli both verbally and visually, which may have improved initial encoding of information. Moreover, the use of visual targets allowed participants to employ spatial cues to assist in initial acquisition and memory. Thus, the VR Office paradigm afforded an environmental advantage over standard instruments, which may have improved performance. While it can be argued that the VR Office test of learning and memory was *easier* than standard neuropsychological measures, the added visual imagery found within the virtual

environments is actually consistent with the rich array of auditory and visual stimulation experienced in the real-world. Additionally, the established neurocognitive literature clearly supports that learning and memory is enhanced when information is provided across multiple sensory modalities (Frick, 1984; Mayer, 1997; Moreno & Mayer, 1999). Thus, while testing within a virtual environment can afford the examinee advantages, these advantages are quite representative of the domain of functioning to which we desire our assessment to generalize.

It might be argued that the similar rate of learning observed between TBI-MET and HC participants is attributable to a ceiling effect. This is a possibility given that the two groups acquired the items within few learning trials ($M= 3.96$). Ceiling effect is unlikely, however, considering that there was substantial range in the number trials required to meet the learning criterion (range, 2 -10 trials). Ceiling effect also is unlikely given that 20% of TBI participants were unable to meet the learning criterion, which suggests the measure was sensitive to memory impairment and was able to discriminate individuals on this basis.

Despite unanticipated differences in initial acquisition of information, participant recall of information generally followed an expected pattern, with an overall greater number of items recalled and recognized after 30 minutes than after 24 hours and TBI-NOT MET participants recalling and recognizing the fewest items after a 30-minute delay. Notably, after 24 hours there were no significant group differences in target recall, where the more impaired TBI-NOT MET group recalled and recognized the same number of items as the other study groups. One potential reason for this is that the TBI-NOT MET group appeared to have lost a large proportion of learned information quickly

(i.e., after 30 minutes), but maintained this reduced set over time (i.e., 24 hours).

Conversely, the TBI-MET and HC groups appeared to have had a slower, more gradual decay of information, which allowed them to demonstrate superior recall after a 30-minute delay. However, continued decay of information over time may explain their similar performance to the TBI-NOT MET group after 24 hours.

While the VR Office differs substantially from standard tests of memory, it was anticipated that it would correlate significantly with an established measure of memory to provide preliminary evidence to support its construct validity. Indeed, there was a significant relationship between the VR Office 30-minute Recall and the CVLT Long Delay-Free Recall indices in which 49% of variance was shared. Clearly, this significant correlation represents only preliminary evidence of construct validity and additional investigation of the psychometric properties of the VR Office as a measure of learning and memory should be the subject of future study. However, based on available data it appears that the VR Office may be tapping memory abilities similar to the established CVLT, but that the two instruments clearly also yield unique data that together may form a more accurate picture of an examinee's real-world performance.

The present study, while adding to the VR assessment literature, was limited by a small sample size, a focused examination of only moderate to severely impaired patients, and a relatively simplified VR environment. Future studies might employ more complicated environments and/or add cognitive tasks that even more closely resemble external activities in an effort to predict real-world performance. Additionally, future studies might increase sensory stimulation with tactile/auditory environmental information in order to further improve ecological validity and to minimize incongruity

between the testing environment and the real-world environment to which we wish to generalize.

Given the obtained results, the VR Office was shown to be a promising means for examining post-TBI learning and memory. While TBI participants' performance in the VR Office did vary slightly from expectations, this may indicate that the environment yields data that is not typically obtained upon traditional neuropsychological evaluation. In specific, whereas standard assessment uses multiple tests to isolate areas of cognitive strength and weakness that then are combined to predict real-world performance, VR assessment may condense this process by assessing more directly a real-world scenario that inherently calls upon an integrated set of skills and cognitive abilities. Our results demonstrated that data regarding integrated memory abilities can be obtained empirically, within a consistent and controllable environment, that resembles the real-world and that maintains the safety of the examinee. Thus, the future of neuropsychological evaluation might employ both the use of our established measures of cognitive function along with VR-based measures in order to fully assess neurocognitive strengths and weakness and provide the most reliable assessment of real-world function following TBI.

ACKNOWLEDGEMENTS

1. This work was supported in part by the Henry H. Kessler Foundation.
2. Our appreciation to NeuroRehabilitation Specialists, Haledon, NJ for assistance in the recruitment of study participants.

REFERENCES

- Benedict, R. H. B., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, M. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8*(2), 145-153.
- Bennett, T. L. (2001). Neuropsychological evaluation in rehabilitation planning and evaluation of functional skills. *Archives of Clinical Neuropsychology, 16*(3), 237-253.
- Bowman, M. L. (1996). Ecological validity of neuropsychological and other predictors following head injury. *The Clinical Neuropsychologist, 10*(4), 382-396.
- Brooks, B. M., Rose, F. D., Potter, J., Jayawardena, S., & Morling, A. (2004). Assessing stroke patients' prospective memory using virtual reality. *Brain Injury, 18*(4), 391-401.
- Buckwalter, J. G., & Rizzo, A. A. (1997). Virtual reality and the neuropsychological assessment of persons with neurologically based cognitive impairments. *Studies in Health Technology and Informatics, 39*, 17-21.
- Crosson, B., Novack, T. A., Trennery, M. R., & Craig, P. L. (1988). California Verbal Learning Test (CVLT) performance in severely head-injured and neurologically normal adult males. *Journal of Clinical and Experimental Neuropsychology, 10*(6), 754-768.

- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version Manual*. San Antonio, TX: The Psychological Corporation.
- DeLuca, J., Schultheis, M. T., Madigan, N., Christodoulou, C., & Averill, A. (2000). Acquisition versus Retrieval Deficits in Traumatic Brain Injury: Implications for Memory Rehabilitation. *Archives of Physical Medicine and Rehabilitation, 81*, 1327-1333.
- Dikmen, S. S., Machamer, J. E., Powell, J. M., & Temkin, N. R. (2003). Outcome 3 to 5 years after moderate to severe traumatic brain injury. *Archives of Physical Medicine and Rehabilitation, 84*, 1449-1457.
- Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., & Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin Card Sorting Test: An analysis of preliminary results. *CyberPsychology and Behavior, 4*(4), 489-496.
- Farias, S. T., Harrell, E., Neumann, C., & Houtz, A. (2003). The relationship between neuropsychological performance and daily functioning in individuals with Alzheimer's disease: ecological validity of neuropsychological tests. *Archives of Clinical Neuropsychology, 18* (6), 655-672.
- Frick, R. W. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory and Cognition, 12*, 507-514.
- Gioia, G. A., & Isquith, P. K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology, 25*(1-2), 135-158.

- Heaton, R. K. (1981). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Hooper, H. E. (1983). *Hooper Visual Organization Test (VOT)*. Los Angeles: Western Psychological Services.
- Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea & Febiger.
- Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology*, *13*(6), 523-534.
- Ku, J., Cho, W., Kim, J. J., Peled, A., Wiederhold, B. K., Wiederhold, M. D., Kim, I. Y., Lee, J. H., & Kim, S. I. (2003). A virtual environment for investigating Schizophrenic patients' characteristics: Assessment of cognitive and navigation ability. *CyberPsychology and Behavior*, *6*(4), 397-404.
- Lezak, M. D. (1995). *Neuropsychological Assessment*. New York: Oxford University Press.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, *32*, 1-19.
- Moreno, R. Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational psychology*, *91*(2). 358-368.

Newnan, O. S., Heaton, R. K., & Lehman, R. A. (1978). Neuropsychological and MMPI correlates of patients' future employment characteristics. *Perceptual Motor Skills, 46*(2), 635-642.

Ptok, M., Buller, N., Kuske, S., & Hecker, H. (2004). Ecological validity of auditory verbal learning and memory tests in children. *Hals-Nasen-Ohren-Heilkunde, 10*.

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting and Clinical Psychology, 19*, 393-394.

Riva, G. (1997). Virtual reality as assessment tool in psychology. *Studies in Health Technology and Informatics, 44*, 71-79.

Rizzo, A. A., & Buckwalter, J. G. (1997). Virtual reality and cognitive assessment and rehabilitation: the state of the art. *Studies in Health Technology and Informatics, 44*, 123-145.

Sbordone, R. J., & Long, C. J. (1996). *Ecological Validity of Neuropsychological Testing*. Delray Beach, FL: St. Lucie Press, Inc.

Schultheis, M. T., Himmelstein, J., & Rizzo, A. R. (2002). Virtual reality and neuropsychology: Upgrading the current tools. *Journal of Head Trauma Rehabilitation, 17*(5), 379-394.

Schultheis, M. T., & Rizzo, A. A. (2001). The application of virtual reality technology in rehabilitation. *Rehabilitation Psychology, 46*(3), 296-311.

- Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests (2nd edition)*. New York: Oxford.
- Vakil, E., Jaffe, R., Eluze, S., Groswasser, Z., & Aberbuck, S. (1996). Word recall vresus reading speed: Evidence of preserved priming in head-injured patients. *Brain and Cognition, 31*, 75-89.
- Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe brain injury. *Neuropsychology, Development, and Cognition: Section C Child Neuropsychology, 8(4)*, 296-303.
- Wechsler. (1999). *Wechsler Abbreviated Scale of Intelligence manual*. New York: Harcourt Brace Jovanovich.
- Wechsler. (1997). *Wechsler Adult Intelligence Scale-Third Edition Administration Manual*. New York: Psychological Corporation.
- Zhang, L., Abreu, B. C., Masel, B., Scheibel, R. S., Christiansen, C. H., Huddleston, N., & Ottenbacher, K. J. (2001). Virtual reality in the assessment of selected cognitive function after brain injury. *American Journal of Physical Medicine and Rehabilitation, 80*, 597-604.
- Zhang, L., Abreu, B. C., Seale, G. S., Masel, B., Christiansen, C. H., & Ottenbacher, K. J. (2003). A virtual reality environment for evaluation of a daily living skill in brain injury rehabilitation: reliability and validity. *American Journal of Physical Medicine and Rehabilitation, 84*, 1118-1124.

Table 1

Demographic Characteristics Organized By Group

	TBI-MET	TBI-NOT MET	HC	χ^2 / F
	(<i>n</i> = 16)	(<i>n</i> = 4)	(<i>n</i> = 20)	
Gender				
Women	7 (44%)	2 (50%)	11 (55%)	0.45
Men	9 (56%)	2 (50%)	9 (45%)	-
Age (in years)				
Mean (<i>SD</i>)	37.6 (9.1)	39.5 (11.8)	31.6 (12.4)	1.71
Education (in years)				
Mean (<i>SD</i>)	13.3 (2.1)	14.0 (2.8)	15.2 (2.4)	3.13
Marital Status				
Single	13 (81%)	3 (75%)	13 (65%)	1.19
Married	3 (19%)	1 (25%)	7 (35%)	
Ethnicity				
Caucasian	15 (94%)	3 (75%)	18 (90%)	10.38*
African American	0	1 (25%)	1 (5%)	
Hispanic	1 (6%)	0	1 (5%)	
Employment Status				
Employed	10 (63%)	2 (50%)	13 (65%)	0.32
Unemployed	0	0	7 (35%)	
Post-TBI Disability	6 (37%)	2 (50%)	--	
Time LOC - days	12.7 (15.3)	29.5 (20.5)	--	3.4

Time Post-Injury (yr.)	7.5 (5.4)	6.5 (4.2)	--	0.12
------------------------	-----------	-----------	----	------

Note. LOC = Loss of consciousness. * $p < 0.05$

Table 2

Summary of Performance on Neuropsychological Measures Among Groups

Measures	TBI-MET		TBI-NOT MET		HC		<i>F</i>
	<i>(n = 16)</i>		<i>(n = 4)</i>		<i>(n = 20)</i>		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
WASI (Age Scaled)							
Full Scale IQ	88.2 ^c	15.2	71.3 ^b	12.2	104.1	14.5	10.8 [*]
Performance IQ	100.4	21.7	84.8	22.7	106.6	16.8	2.2
Vocabulary	5.3 ^c	4.1	1.0 ^b	0.0	10.5	4.6	12.0 [*]
Block Design	9.1	3.5	6.8	3.9	10.7	4.4	1.8
Matrix Reasoning	11.4	3.3	7.5	5.1	12.4	4.1	2.6
WAIS-III (Age Scaled)							
Digit Symbol Coding	6.5 ^c	2.3	4.8 ^b	2.6	10.5	2.6	15.4 [*]
Digit Span Total	9.3	4.1	6.3 ^b	1.9	11.8	2.4	6.2 ^{**}
CVLT (Items Recalled)							
Total Trials 1-5	44.8 ^c	9.4	25.0 ^{ab}	4.7	55.9	8.1	24.7 [*]
Trial B	4.8 ^c	2.4	3.5 ^b	1.3	7.0	2.1	6.8 ^{**}
Short Delay Free Recall	9.2 ^c	3.1	3.3 ^{ab}	2.8	11.9	3.3	12.9 [*]
Long Delay Free Recall	8.9 ^c	3.5	2.3 ^{ab}	2.2	12.8	3.0	20.9 [*]
BVMT (Details Recalled)							
Total Trials 1-3	20.5 ^c	7.1	12.8 ^b	7.6	28.4	6.3	11.9 [*]
30min. Delay	7.7 ^c	3.3	3.8 ^{ab}	2.2	11.0	1.6	18.0 [*]

Trail Making Test (sec.)							
Part A	42.4	19.5	88.3 ^{ab}	62.6	27.5	9.0	12.2 [*]
Part B	105.6 ^c	58.1	162.8 ^b	12.01	57.8	20.6	8.2 ^{**}
Wisconsin Card Sort Test							
Categories Completed	4.3	2.4	2.3 ^b	2.9	5.5	1.7	4.4 ^{**}
Perseverative Errors	15.8	13.0	20.0	16.6	9.7	7.8	2.2
Boston Naming Test	26.9	4.3	22.3 ^{ab}	5.5	29.1	1.7	7.3 ^{**}
HVOT (Total Correct)	25.0	4.1	20.0 ^b	2.5	26.3	4.0	4.3 ^{**}
COWAT (Total)	30.4 ^c	11.3	26.3 ^b	14.7	44.3	12.1	7.7 ^{**}

Note. ^{*} $p < 0.001$; ^{**} $p < 0.01$; ^a Significant difference between TBI-Met and TBI-Not Met groups ($p < 0.01$); ^b Significant difference between TBI-Not Met and HC groups ($p < 0.01$). ^c Significant difference between TBI-MET and HC groups ($p < 0.01$).

Figure 1



Figure 2

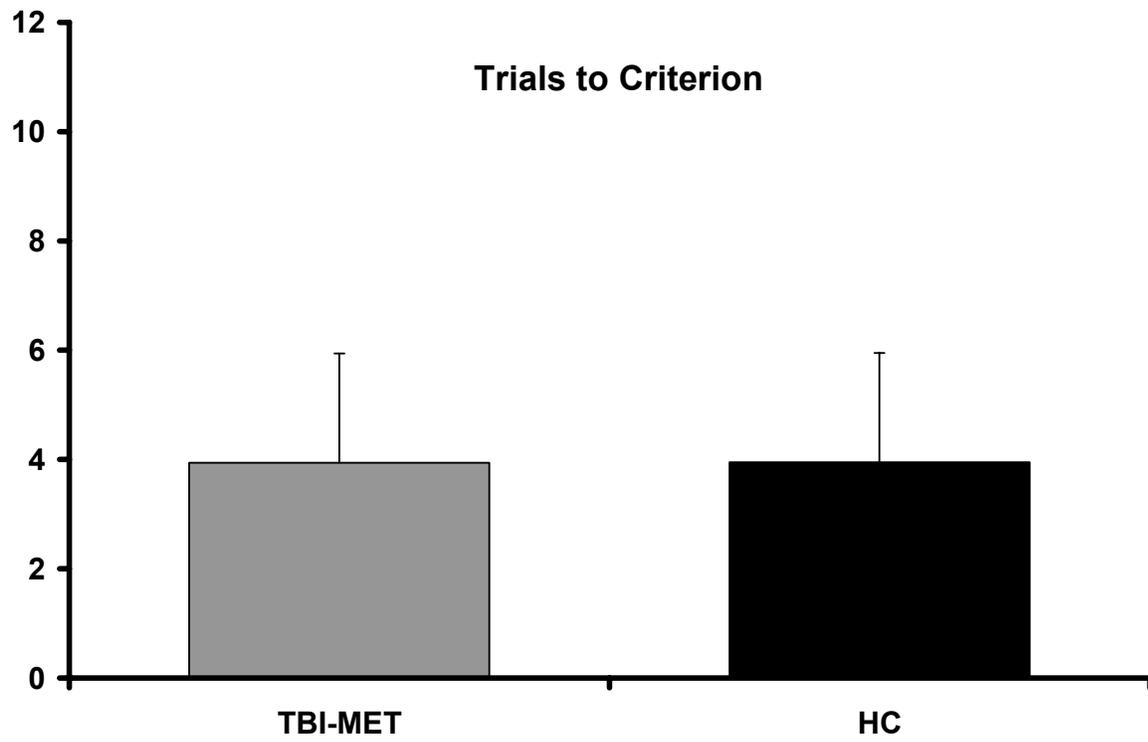


Figure 3

