Neuropsychological assessment using the virtual reality cognitive performance assessment test

T D Parsons and A A Rizzo

Institute for Creative Technologies, University of Southern California,
Los Angeles, California, USA

tparsons@usc.edu, *arizzo@usc.edu*

http://vrpsych.ict.usc.edu/

ABSTRACT

The traditional approach to assessing neurocognitive performance makes use of paper and pencil neuropsychological assessments. This received approach has been criticized as limited in the area of ecological validity. The newly developed Virtual Reality Cognitive Performance Assessment Test (VRCPAT) focuses upon enhanced ecological validity using virtual environment scenarios to assess neurocognitive processing. The VRCPAT battery and a neuropsychological assessment were conducted with a sample of healthy adults. Findings suggest 1) good construct validity for the Memory Module; and 2) that increase in stimulus complexity and stimulus intensity can manipulate attention performance within the Attention Module.

1. INTRODUCTION

While standard neuropsychological measures have been found to have adequate predictive value, their ecological validity may diminish predictions about real world functioning (Chaytor et al., 2006; Farias, Harrell, Neumann, & Houtz, 2003; Gioia & Isquith, 2004; Odhuba et al., 2005). Traditional neurocognitive measures may not replicate the diverse environment that in which persons live. Additionally, standard neurocognitive batteries tend to examine isolated components of neuropsychological ability, which may not accurately reflect distinct cognitive domains (Parsons et al., 2005). Virtual Reality (VR) technology is increasingly being recognized as a useful tool for the study, assessment, and rehabilitation of cognitive processes and functional abilities. The ability of VR to create dynamic, immersive, three-dimensional stimulus environments, in which all behavioral responding can be recorded, offers assessment and rehabilitation options that are not available using traditional assessment methods. In this regard, VR applications are now being developed and tested which focus on component cognitive processes including: attention processes (Parsons et al., in press; Rizzo et al., 2006), spatial abilities (Parsons et al., 2004), memory (Matheis et al., 2007), and executive functions (Baumgartner et al., 2006; Elkind et al., 2001). The increased ecological validity of neurocognitive batteries that include assessment using VR scenarios may aid differential diagnosis and treatment planning.

The Virtual Reality Cognitive Performance Assessment Test (VRCPAT) project focuses on the refined analysis of neurocognitive testing using a virtual environment to assess recall of targets delivered within the context of a virtual city. Herein we report on two phases of VRCPAT development: 1) the psychometric properties of data gained from human pilot testing with the VRCPAT; and 2) Attentional processing within the virtual environment.

2. METHODS: EXPERIMENT 1

We acquired data on the implementation of a virtual reality (i.e. VRCPAT) in a normative sample that also received a traditional paper and pencil battery. Because the VRCPAT was designed to tap very specific neurocognitive systems and not to mirror a traditional paper-and-pencil battery, our goal is not to replace the traditional battery for all neurocognitive domains. We aim to assess the psychometric properties of the VR and paper-and-pencil measures. Hence, scores were correlated with demographic and other performance tests measures administered. Standard correlational analyses using a brief demographic survey and pencil-and-paper cognitive tests aid our initial assessment of both the concurrent and divergent validity properties of this form of assessment.

Our plan for the development and implementation of the VRCPAT's psychometric properties involved systematic refinement analyses that acted as a component of an ongoing dialectic between measurement and substantive research. We aim to make the VRCPAT to be a well developed measure that facilitates substantive advances. First, we identified the VRCPAT's hierarchical or aggregational structure. Next, we established the internal consistency of the VRCPAT's unidimensional facets (memory domain). We also determined the content homogeneity of each of the VRCPAT's unidimensional facets. The establishment of the VRCPAT's psychometric properties removed the possibility that results reflect correlates of the target construct (memory and/or attention) but are not prototypic of it. We also assessed the level to which all aspects of the target construct (memory) is under- or overrepresented in the VRCPAT's composition, and assess whether the experience of some aspects of the virtual environment introduced variance unrelated to the target construct.

2.1 Participants

The study sample included 20 healthy subjects (Age, mean = 24.45, SD = 3.05; 50 % male; and Education, mean = 14.05, SD = 0.51). Strict exclusion criteria were enforced so as to minimize the possible confounding effects of comorbid factors known to adversely impact cognition, including psychiatric (e.g., mental retardation, psychotic disorders, diagnosed learning disabilities, Attention-Deficit/Hyperactivity Disorder, and Bipolar Disorders, as well as substance-related disorders within two years of evaluation) and neurologic (e.g., seizure disorders, closed head injuries with loss of consciousness greater than 15 minutes, and neoplastic diseases) conditions. Subjects were comparable in age, education, ethnicity, sex, and self-reported symptoms of depression.

2.2 Procedure

After providing informed written consent, all participants were administered the VRCPAT as part of a larger neuropsychological test battery. The VRCPAT is a 15-minute measure, in which participants (referred to as "users" in the following text) then go through the following steps: <u>Acquisition Phase</u> – Users are presented with 10 pieces of language-based information to be learned, without any context for what they will need to do with this information. The information stimuli is primarily language based (i.e., blue car with bullet holes in windshield only, intact barrel with U.S. Army stenciled on it, red shipping container with numbers 7668, etc.), although stimuli includes minimal imagery to provide opportunities for more context relevant stimulus creation (e.g., a crate with Iraqi flag in upper side corner (with just image of Iraqi flag presented)). The acquisition phase is initially standardized to three one-minute trials. At the end of each trial, users are then asked to name the objects that they studied as an assessment of initial declarative recall memory. It should be noted here that the item pool of stimuli that is being used in the formal tests was generated during the initial development phase. At that time, various stimulus combinations were piloted (not using VR) with users to determine "memorability" to inform final selection of stimuli that is being used in the formal VR test. Informal exploration of image-based stimuli for later development of a pure object-based visual stimulus MM test has occurred concurrently with verbal tests, using pictures of objects similar to the language stimuli.

2.2.1 VR Interface and Task Training Phase. After users are given the three one-minute trials to "memorize" the stimuli, a brief "interface training" period then occurs in which users become familiar with their objective, the controls of the game pad navigation interface and head-mounted display (HMD). The task is read aloud by the investigator and contains specific instructions for how to proceed through the VE and how to record images of each target object. Users are told that they will need to go to each target zone in sequence (guided by an American soldier), and at each zone, two of the items that they had memorized previously will be present somewhere in the environment from that vantage point. Upon finding the items, they should align the cross hairs with that object and press the response button to record (or "collect" them). Users have one minute to spend within each target zone and scan for the relevant memorized target items. If they find the target items in less than the one-minute period, they must wait for time to expire and follow their guide to the next zone. If the user does not find both objects in a target zone by the time that the one-minute period has elapsed, an alarm sounds and a voice tells the user to move to the next zone and seek out the two objects located there. To minimize guessing by subjects that hit the response button on all possible target objects, subjects are told that they have a limit of two button presses per zone.

The VE is designed to resemble and Iraqi city and the location of this task training phase also serves as the starting point for the users. The environment contains people (of various ages and gender) dressed in culturally appropriate clothing, a market place, American soldiers, various moving civilian vehicles, animals, buildings and street signs and a host of other objects (i.e., wooden barrels, crates, containers, etc.). Users are given as much time as needed to explore a limited area of the environment. This exploration area is determined by the experimenter. During this phase, the investigator can informally present verbal guidance to the users in order to help them to become familiar with the interface navigation, response button and HMD viewing parameters. This phase is designed to teach the interface controls to the user, so that performance on the VE navigation and object selection interaction tasks are minimally influenced or distracted away from the resulting memory assessment in the retrieval phase that follows.

2.2.2 Retrieval Phase. Once users indicate that they are comfortable within the VE and can demonstrate comprehension of the navigation interface and targeting procedure, the investigator asks if there are any questions. If so, clarification and coaching occur until the user can fully comprehend the task. If not, the investigator starts the exercise and observes the user. Should problems occur, the investigator can verbally coach the user until issues are resolved.

2.2.3 Debriefing Phase. During this phase, users are asked to recall the original list of stimuli and at which target zones they were found. The performance measures that are derived from this test include: number of correct hits, false hits, time to successfully complete per target zone, time to complete overall. A trained research assistant administered all psychometric tests. The Simulator Sickness Questionnaire (Kennedy et al., 1992; SSQ) was used to determine whether the participant felt sick as a result of the VR experience.

The broader neuropsychological battery contained the following tests: 1) Controlled Oral Word Association Test (COWAT-FAS; Benton, Hamsher, & Sivan, 1994; Gladsjo et al., 1999); 2) animal fluency (Gladsjo et al., 1999); 3) Digit Symbol Coding, Digit Span (Forward and Backward), and Letter-Number Sequencing from the Wechsler Adult Intelligence Scale –Third edition (WAIS-III; Psychological Corporation, 1997); 4) Trail Making Test Parts A and B (TMT; Heaton, Grant, & Matthews, 1991; Reitan & Wolfson, 1985); 5) Hopkins Verbal Learning Test – Revised (HVLT-R; Brandt & Benedict, 2001); 6) Brief Visuospatial Memory Test – Revised (BVMT-R; Benedict, 1997); 7) Stroop Color and Word Test (Golden, 1978); 8); Wechsler Test of Adult Reading (WTAR; Wechsler, 2001).

3. RESULTS: EXPERIMENT 1

Given the similarity of participants in terms of age, sex, education, immersiveness, and ethnicity, no correction for these variables was employed. Notably, none of the participants reported simulator sickness following VR exposure as measured by the SSQ. To provide preliminary data to support the validity of the VRCPAT as a measure of learning and memory, recall indices from the VRCPAT and traditional neuropsychological tests were correlated. Indices were developed from linear composites derived from z-score transformations. Specifically, Pearson correlation analyses were used to compare recall from the VRCPAT with linear composites derived from traditional neuropsychological measures.

3.1 Convergent Validity Tests

Whilst the VRCPAT Total Memory Score was significantly correlated with composites derived from established measures of learning and memory, it did not correlate with possibly confounded variables (i.e., Executive Functions; Attention; and Processing Speed) drawn from traditional neuropsychological measures that are not assessments of learning and memory. Hence, the results indicated that the VRCPAT correlated significantly with the traditional neuropsychological Learning Composite (HVLT Trials 1-3; and BVMT Trials 1-3; $r = 0.68$, $p < 0.001$), with 46% variance shared between the two indices. The results indicated that the VRCPAT also correlated significantly with the traditional neuropsychological Memory Composite (HVLT Total Recall after a Delay; and BVMT Total Recall after a Delay; $r = 0.67$, $p < 0.001$), with 45% variance shared between the two indices.

3.2 Discriminant Validity Tests

As expected, there were no significant correlations between VRCPAT measures and the following neuropsychology test composites: Executive Functions Composite; Attention Composite; or Processing Speed Composite. Hence, each of the discriminant validity significance tests were as predicted, that is, did not correlate with theoretically unrelated abilities. Although validity coefficients drawn from composites may not meet validity expectations it may still be the case that individual measures account for some of the trait variance. Therefore, we assessed the measures both as composites and individually. As such, we compared the VRCPAT with the actual neuropsychological tests (used to derive the Learning Composite and the Memory Composite). Analysis of the relations between the VRCPAT Total Memory Score and the actual learning and memory tests revealed significant correlations for each of the convergent validity significance tests, in accordance with prediction. For correlations between the VRCPAT and traditional psychometric measures we only considered those correlations that met the criterion of $p < .05$ to be meaningful. Given our small sample size we kept p at this level, despite the risk of Type I error with multiple correlations. All of our significant correlations were associated with at least moderate effect sizes.

4. METHODS: EXPERIMENT 2

The following two VR-based attentional measures were designed and evolved following iterative user testing: 1) Fixed Position in the Virtual City Test (See Figure 1); and 2) Humvee Attention Task.

Figure 1. *Fixed Position in the Virtual City.*

Figure 2. *Humvee Attention Task.*

4.1.1 Fixed Position in the Virtual City Test. In this scenario subjects were given both a selective attention and a working memory task. For the selective attention portion, each subject listened to a virtual trainee as the trainee classified passing vehicles. For the evaluation, the virtual trainee reported either "US military", "Iraqi police", "Iraqi civilian" or "possible insurgent". The subject was to tell the new recruit whether he was correct or incorrect. For the working memory portion, subjects were presented a series of single digit numbers. Subjects listened for the first two numbers, added them up, and reported the answer to the examiner. When the subject heard the next number, s/he added it to the one presented right before it. Subjects continued to add the next number to each preceding one. Subjects were not being asked to give examiner a running total, but rather the sum of the last two numbers that were presented. For example, if the first two numbers were '5' and '7,' subject would say '12.' If the next number were '3,' subject would say '10.' Then if the next number were '2,' subject would say '5' because the last two numbers presented were 3 and 2. See Table 1 for descriptives.

4.1.2 HUMVEE Attention Task. The Humvee scenario assessed attention within both "safe" and "ambush" settings: 1) start section; 2) palm ambush; 3) safe zone; 4) city ambush; 5) safe zone; 6) bridge ambush. The task involved the presentation of a four-digit number that was superimposed on the virtual windshield (of the Humvee) while the subject drove the Humvee. Each four-digit number was presented for approximately 300 ms and was randomly selected by the computer from a database of prescreened numbers. Subjects were required to say the number out loud immediately after it appeared on the screen while the Humvee continued driving. An examiner will recorded the responses. See Table for descriptives of Humvee Attention Test.

The design consists of six Humvee attention conditions:

i) *Fixed Position: 2.0 second condition (Start Section).* In this condition, the four-digit number always appeared in a *fixed central* location on the "windshield." The numbers were presented at 2.0 second intervals. This occurred in the "Start Section" and ended just before the "Palm Ambush."

ii) *Fixed Position: 1.5 second condition (Palm Ambush).* The procedure for this condition was identical to the "Fixed Position" condition described previously except that the numbers were presented at 1.5 second intervals. This occurred in the "Palm Ambush" section and ended just before the "Safe Zone" section.

iii) *Fixed Position: 0.725 second condition (Safe Zone).* The procedure for this condition was identical to the "Fixed Position" condition described previously except that the numbers were presented at 0.725 second intervals. This occurred in the "Safe zone" and ended just before the "City Ambush" section.

iv) *Random Position: 2.0 second condition (City Ambush).* The procedure for this condition is similar to the "Fixed Position" condition with the exception that the numbers appear *randomly* throughout the "windshield" rather than in one fixed central location. The numbers were presented at 2.0 second intervals. This occurred in the "City Ambush" and ended just before the "Safe Zone".

v) Random Position: 1.5 second condition (Safe Zone). The procedure for this condition is similar to the preceding "Random Position" condition except that the numbers were presented at 1.5 second intervals. This occurred in the "Safe Zone" and ended just before the "Bridge Ambush".

vi) Random Position: 0.725 second condition (Bridge Ambush). The procedure for this condition is similar to the preceding "Random Position" condition except that the numbers were presented at 0.725 second intervals. This occurred in the "Bridge Ambush".

5. RESULTS: EXPERIMENT 2

To examine scenario differences, one-way ANOVAs were performed, comparing attentional performance in simple stimulus presentations (Mean = 43.63; SD = 8.91) versus complex stimulus presentations (Mean = 34.63; SD = 6.86). The results indicated that the increase in stimulus complexity caused a significant decrease in performance on attentional tasks ($F = 5.12$; $p = 0.04$). To examine scenario differences, we compared attentional performance in low intensity (Mean = 40.01; SD = 4.06) versus high intensity (Mean = 9.25; SD = 3.70) presentations. The results indicated that the increase in stimulus intensity caused a significant decrease in performance on attentional tasks ($t = 9.83$; $p = 0.01$). Given the small sample size, we decided to not assess the construct validity of the VRCPAT Attention Modules. Hence, no attempts were made to assess correlations between standard paper and pencil tests and VRCPAT.

6. DISCUSSION

6.1 Experiment 1

The results of this study indicate that: 1) VRCPAT memory measures correlated significantly with scores from the memory measures drawn from the traditional neuropsychological test battery; 2) VRCPAT memory scores did not correlate with non-memory measures drawn from the traditional neuropsychological test battery. Additionally, no negative side effects were associated with use of the VRCPAT. The establishment that the VRCPAT's memory measures correlated significantly with scores from the memory measures drawn from the traditional neuropsychological test battery but not with non-memory measures removed the possibility that results reflected correlates of the non-target construct (i.e. processing speed; executive function).

Our goal was to conduct an initial pilot study of a VRCPAT that employs a standard neuropsychological battery for the assessment of normal participants. We believe that this goal was met. We recognize, however, that the current findings are only a first step in the development this tool. Many more steps need to be taken in order to continue the process of test development and to fully establish the VRCPAT as a measure that contributes to existing assessment procedures for the diagnosis of memory decline. Whilst the VRCPAT as a measure needs to be fully validated, current findings provide preliminary data regarding the validity of the virtual environment as a memory measure. The VRCPAT was correlated with widely used memory assessment tools. Nevertheless, the fairly small sample size requires that the reliability and validity of the VRCPAT be established using a larger sample of well-matched participants. This will ensure that current findings are not a sample size related anomaly. Finally, the ability of the VRCPAT to accurately classify participants not involved in the initial validation study will need to be examined for cross-validation purposes.

6.2 Experiment 2

Our goal was to conduct an initial pilot study of the general usability of the VRCPAT Attention Module scenarios. We aimed at assessing whether the increase in stimulus complexity would result in a significant decrease in performance on attentional tasks. We also wanted to see whether an increase in stimulus intensity would result in a significant decrease in performance on attentional tasks. We believe that this goal was met as the study results indicated that: (1) the increase in stimulus complexity caused a significant decrease in performance on attentional tasks; and 2) the increase in stimulus intensity caused a significant decrease in performance on attentional tasks.

Our findings should be understood in the context of some limitations. First, these findings are based on a fairly small sample size. As a necessary next step, the reliability and validity of the test needs to be established using a larger sample of participants. This will ensure that the current findings are not an anomaly due to sample size. Additionally, the diagnostic utility of this attention assessment tool must be determined. The ability of the VRCPAT's Attention Module to accurately classify participants into attention impaired and attention intact groups based on carefully established critical values must be evaluated. This will involve the generation of specific cut-off points for classifying a positive or negative finding. The VRCPAT Attention

Module's prediction of attentional deficits will need to be evaluated by the performance indices of sensitivity, specificity, predictive value of a positive test, and predictive value of a negative test.

In sum, manipulation of stimulus complexity and intensity in the VRCPAT's Attention Module caused a significant differences in performance on attentional tasks. Complementary comparisons of the VRCPAT's Attention Module with behavioral and neurocognitive tests developed to assess attentional processing are also warranted to determine the construct validity of the test.

7. REFERENCES

T Baumgartner, L Valko, M Esslen and L Jancke (2006), Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study, *Cyberpsychology & Behavior, 9*, 30–45.

R H B Benedict (1997), *Brief Visuospatial Memory Test-revised*, Odessa, FL: Psychological Assessment Resources, Inc.

L A Benton, K Hamsher and A B Sivan (1994), Controlled Oral Word Association Test, In *Multilingual aphasia examination* (3rd ed.), Iowa City, IA: AJA.

J Brandt and R H B Benedict (2001), *Hopkins Verbal Learning Test—revised, Professional manual*, Lutz, FL: Psychological Assessment Resources, Inc.

N Chaytor, M Schmitter-Edgecombe and R Burr (2006), Improving the ecological validity of executive functioning assessment, *Archives of Clinical Neuropsychology, 21*, 217–227.

J S Elkind, E Rubin, S Rosenthal, B Skoff and P Prather (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.

S T Farias, E Harrell, C Neumann and A Houtz (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.

G A Gioia and P K Isquith (2004), Ecological assessment of executive function in traumatic brain injury, *Developmental Neuropsychology, 25(1-2)*, 135-158.

J A Gladsjo, C C Schuman, J D Evans, G M Peavy, S W Miller and R K Heaton (1999), Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity, *Assessment, 6*, 147–178.

C J Golden (1978), *Stroop Color and Word Test*, Chicago, IL: Stoelting.

R K Heaton, I Grant and C G Matthews (1991), *Comprehensive norms for an expanded Halstead- Reitan Battery: Demographic corrections, research findings, and clinical applications*, Odessa, FL: Psychological Assessment Resources.

R S Kennedy, J E Fowlkes, K S Berbaum and M G Lilienthal (1992), Use of a motion sickness history questionnaire for prediction of simulator sickness, *Aviation, Space & Environ. Medicine, 63,* 588–593.

R J Matheis, M T Schultheis, L A Tiersky, J Deluca, S R Millis and A A Rizzo (2007) Is learning and memory different in a virtual environment? *Clin Neuropsychol. 2007 Jan;21(1)*, 146-61.

R A Odhuba, M D van den Broek and L C Johns (2005), Ecological validity of measures of executive functioning, *The British Journal of Clinical Psychology, 44*, 269–278.

T D Parsons, T Bowerly, J G Buckwalter and A A Rizzo (2007), A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods, *Child Neuropsychology*.

T D Parsons, P Larson, K Kratz, M Thiebaux, B Bluestein, J G Buckwalter et al (2004), Sex differences in mental rotation and spatial rotation in a virtual environment, *Neuropsychologia, 42*, 555–562.

T D Parsons, A A Rizzo, C van der Zaag, J S McGee and J G Buckwalter (2005), Gender and Cognitive Performance: A Test of the Common Cause Hypothesis, *Aging, Neuropsych. & Cognition, 12(1)*, 78 - 88.

R M Reitan and D Wolfson (1985), *The Halstead-Reitan Neuropsychological test battery: Theory and clinical interpretation,* Tucson, AZ: Neuropsychology Press.

A A Rizzo, T Bowerly, J G Buckwalter, D Klimchuk, R Mitura and T D Parsons (2006), A Virtual Reality Scenario for All Seasons: The Virtual Classroom, *CNS Spectrums, 11(1)*, 35-44.

D Wechsler (1997), *Wechsler Adult Intelligence Scale – Third edition*, San Antonio, TX: Psychological Corporation.

D Wechsler (2001), *Wechsler Test of Adult Reading (WTAR)*, New York: NY: Psychological Corporation.