

Virtual Reality Cognitive Performance Assessment Test

Thomas D. Parsons, Albert A. Rizzo, Jacob Bamattre, & John Brennan
University of Southern California, Institute for Creative Technologies

Abstract

Virtual Reality Cognitive Performance Assessment Test (VRCPAT) is a virtual environment-based measure of learning and memory. We examined convergent and discriminant validity and hypothesized that the VRCPAT's Total Learning and Memory scores would correlate with other neuropsychological measures involving learning and memory, but not with measures involving potential confounds (i.e., Executive Functions; Attention; and Processing Speed). Using a sequential hierarchical strategy, each stage of test development did not proceed until specified criteria were met. The 15 minute VRCPAT battery and a 1.5 hour in-person neuropsychological assessment were conducted with a randomly selected sample of 20 healthy adults that included equivalent distributions of men and women from ethnically diverse populations. Results supported both convergent and discriminant validity. That is, findings suggest that the VRCPAT measures a capacity that is 1) consistent with that assessed by traditional paper and pencil measures involving learning and memory; and 2) inconsistent with that assessed by traditional paper and pencil measures assessing neurocognitive domains traditionally assumed to be other than learning and memory. We conclude that the VRCPAT is a valid test that provides a unique opportunity to reliably and efficiently study memory function within an ecologically valid environment.

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; Othuba et al., 2005). Traditional neurocognitive measures may not replicate the diverse environment 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 that 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. Following the general psychometric conventions, any measure purporting to quantify a particular neurocognitive domain should be highly correlated with other measures of the same neurocognitive domain (convergent validity), whereas it should not be too highly correlated with tests of different neurocognitive domains (discriminant validity). Herein we report on the psychometric properties of data gained from human pilot testing with the VRCPAT.

Methods

We acquired data on the implementation of a virtual reality battery (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 test 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 a well-developed measure that facilitates substantive advances. We 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.

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.

Procedure: The University of Southern California's Institutional Review Board approved the study. Experimental sessions took place over a two hour period. After informed consent was obtained, basic demographic information and computer experience and usage activities were recorded. Subjects then completed a neuropsychological battery administered under standard conditions. Following completion of the neuropsychological battery, subjects completed the simulator sickness questionnaire (Kennedy, Lande, Berbaum, & Lilienthal, 1992), which

includes a pre-VR exposure symptom checklist. Next, all participants were administered the VRCPAT as part of a larger neuropsychological test battery.

The following traditionally used paper and pencil neuropsychological measures were used as convergent validity measures because each is considered to have an important memory component, and has been used clinically to estimate memory abilities: To assess verbal learning and memory we used the Hopkins Verbal Learning Test – Revised (HVLTR; Brandt & Benedict, 2001); to assess nonverbal learning and memory we used the Brief Visuospatial Memory Test – Revised (BVMT-R; Benedict, 1997); and to assess Lexical-Semantic Memory we used Controlled Oral Word Association Test (FAS; Benton, Hamsher, & Sivan, 1994); 2) Semantic Fluency (Animals; Gladsjo et al., 1999);

Discriminant validity measures that were drawn from the corpus of traditionally used paper and pencil neuropsychological measures included: to assess Attention we used Digit Span (Forward and Backward) from the Wechsler Adult Intelligence Scale –Third edition (WAIS-III; Psychological Corporation, 1997); to assess processing speed we used Digit Symbol Coding from the Wechsler Adult Intelligence Scale –Third edition (WAIS-III; Psychological Corporation, 1997), and Trail Making Test Part A (TMT; Heaton, Grant, & Matthews, 1991; Reitan & Wolfson, 1985); to assess executive functioning we used Trail Making Test Part B (TMT; Heaton, Grant, & Matthews, 1991; Reitan & Wolfson, 1985) and the Stroop Color and Word Test (Golden, 1978).

The VRCPAT is a 15-minute measure, in which participants (referred to as “users” in the following text) then go through the following steps: Acquisition Phase – 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, intact barrel with stenciling on it, red shipping container with numbers 7668, etc.), although stimuli includes minimal imagery to provide opportunities for more context relevant stimulus creation. 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 test has occurred concurrently with verbal tests, using pictures of objects similar to the language stimuli.

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 virtual environment (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, 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 a 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, avatars, various moving and parked 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.

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.

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.

Data Analytics

All data were analyzed using SAS version 9.1 (SAS Institute Inc, Cary, NC, 2003). Descriptive statistics were calculated for participant demographics, and for results of the VRCPAT and the criterion neuropsychological tests. Missing data were imputed by either “mean substitution” or “last case carried forward.” Two types of analyses were performed. The first involved computing basic correlations between VRCPAT measures and traditional neuropsychological measures assessing learning and memory. Two-tailed Pearson correlations were used. To make the results for continuous tests comparable, the test scores were converted to standard scores with a mean of zero and standard deviation of one (z-score transformation). The second series of analyses involved computing basic correlations between VRCPAT measures and traditional neuropsychological measures that assess domains that have traditionally been understood to be other than learning and memory. Two-tailed Pearson correlations were used. To make the results

for continuous tests comparable, the test scores were converted to standard scores with a mean of zero and standard deviation of one (z-score transformation). It was expected that while VRCPAT Learning and Memory scores would be highly correlated with standard neuropsychological measures of learning and memory, VRCPAT Learning and Memory scores would not be significantly correlated with neuropsychological measures of domains other than memory.

Results

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. Table 1 shows correlations between traditional test cognitive domain scores and VRCPAT scores.

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.

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 (see Table 1). Hence, the results indicated that the VRCPAT correlated significantly with the traditional neuropsychological Learning Composite (HVLТ 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 (HVLТ 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.

Table 1: Correlations between traditional test cognitive domain scores and VRCPAT scores

	VR Learning		VR Recognition		VR Speed		VR Memory	
	R	P	R	p	r	p	R	p
Learning	0.68	*	0.72	*	0.39	*	0.70	*
Retrospective Memory	0.69	*	0.67	*	0.29		0.65	*
Executive Functions	-0.14		-0.25		0.19		-0.05	
Attention	0.25		0.25		0.16		0.18	
Processing Speed	-0.17		-0.19		-0.01		-0.20	
Lexical-Semantic Memory	0.67	*	0.31		0.13		0.51	*

Note: N=20 for all analyses. VR = Virtual Reality.

* $P < 0.05$

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 (Table 2).

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 (Table 2).

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.

Table 2: Correlations between traditional neurocognitive test scores and VRCPAT total memory score

	R	* P
Learning		
HVLT Trials 1–3	0.59	< 0.05
BVMT-R Trials 1–3	0.43	< 0.05
Retrospective Memory		
HVLT Recognition	0.57	< 0.05
BVMT-R Recognition	0.36	< 0.05
Executive Functions		
Trails B	-0.10	ns
Stroop Interference	0.01	ns
Attention		
Letter Number Sequencing	0.21	ns
Digit Span Forward	0.10	ns
Digit Span Backward	0.10	ns
Processing Speed		
Trails A	-0.13	ns
Digit Coding	-0.09	ns
Lexical-Semantic Memory		
Animals	0.46	< 0.05
Letter Fluency	0.27	$p=.171$

Note: N=20 for all analyses.

Discussion

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 findings should be understood in the context of some limitations. 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, as indicated above, the diagnostic utility of this VRCPAT assessment tool must be determined. The ability of the VRCPAT to accurately classify participants into memory impaired and non-impaired groups based on carefully established critical values must be evaluated. This will involve the generation of specific cut-off points for classifying a positive (memory impaired likely) or negative (memory impaired unlikely) finding. The VRCPAT's prediction of memory impairment 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. Even though reliability is considered to be a unique asset of testing in computer-generated VEs, issues of test-retest reliability need to be addressed.

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 of 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.

References

- Baumgartner, T., Valko, L., Esslen, M., & Jancke, L. (2006). Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study. *Cyberpsychology & Behavior, 9*, 30–45.
- Benedict, R.H.B. (1997). *Brief Visuospatial Memory Test-revised*. Odessa, FL: Psychological Assessment Resources, Inc.

- Benton, L.A., Hamsher, K., & Sivan, A.B. (1994). Controlled Oral Word Association Test. In *Multilingual aphasia examination* (3rd ed.). Iowa City, IA: AJA.
- Brandt, J., & Benedict, R.H.B. (2001). *Hopkins Verbal Learning Test—revised. Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*, 217–227.
- 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.
- Gioia, G.A., & Isquith, P.K. (2004). Ecological assessment of executive function in traumatic brain injury. *Developmental Neuropsychology*, *25*(1-2), 135-158.
- Gladsjo, J.A., Schuman, C.C., Evans, J.D., Peavy, G.M., Miller, S.W., & Heaton, R.K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, *6*, 147–178.
- Golden, C. J. (1978). *Stroop Color and Word Test*. Chicago, IL: Stoelting.
- Heaton, R.K., Grant, I., & Matthews, C.G. (1991). *Comprehensive norms for an expanded Halstead- Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.
- Kennedy, R. S., Fowlkes, J. E., Berbaum, K. S., & Lilienthal, M. G. (1992). Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, and Environmental Medicine*, *63*, 588–593.
- Matheis, R.J., Schultheis, M.T., Tiersky, L.A., Deluca, J., Millis, S.R., & Rizzo, A. Is learning and memory different in a virtual environment? *Clin Neuropsychol*. 2007 Jan;21(1), 146-61.
- Odhuba, R.A., van den Broek, M.D., & Johns, L.C. (2005). Ecological validity of measures of executive functioning. *The British Journal of Clinical Psychology*, *44*, 269–278.
- Parsons, T.D., Bowerly, T., Buckwalter, J.G., & Rizzo, A.A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*.
- Parsons, T.D., Larson, P., Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J.G., et al. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, *42*, 555–562.
- Parsons, T.D., Rizzo, A.A., van der Zaag, C., McGee, J.S., & Buckwalter, J.G. (2005). Gender and Cognitive Performance: A Test of the Common Cause Hypothesis. *Aging, Neuropsychology, and Cognition*, *12*(1), 78 - 88.
- Reitan, R.M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological test battery: Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press.
- Rizzo, A., Bowerly, T., Buckwalter, J., Klimchuk, D., Mitura, R., Parsons, T.D. (2006). A Virtual Reality Scenario for All Seasons: The Virtual Classroom. *CNS Spectrums*, *11*(1), 35-44.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third edition*. San Antonio, TX: Psychological Corporation.

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