An Initial Validation of Virtual Human Administered Neuropsychological Assessments

Thomas D. PARSONSa,*, Paul SCHERMERHORNb, Timothy MCMANHA, Justin ASBEEa, and Nicole RUSSOA

aComputational Neuropsychology and Simulation (CNS) Lab, Department of Psychology, University of North Texas, Denton, Texas USA
bSoartech, Ann Arbor Michigan USA

Abstract. Appropriate neuropsychological assessments are often not readily accessible due to remote geographical locations. The current study performs construct validation of a virtual human-administered neuropsychological assessment of learning and memory. Construct validity was assessed through comparison of virtual human with face-to-face (human) administered neuropsychological assessments. Computerized (virtual human) learning and memory scores were hypothesized to correlate with traditional measures of learning and memory, but not with tests of executive functions or picture naming. Convergent validity results indicated that AVANT learning and memory tasks correlated significantly with traditional learning and memory tests. Divergent validity was found in that AVANT learning was not correlated with traditional executive functioning or confrontational word retrieval measures. Likewise, no significant correlations were found between AVANT memory and executive function or confrontational word retrieval. Findings suggest that the virtual human tests measure a capacity that is consistent with that assessed by traditional paper-and-pencil measures involving learning and memory; and inconsistent with traditional measures of domains other than learning and memory. We conclude that the virtual human-based computerized assessment is a valid test of learning and memory.

Keywords. Virtual Reality; Neuropsychology; Virtual Human; Psychometrics

1. Introduction

Neurocognitive dysfunction adversely affects people’s ability to perform the complex cognitive and affective processing tasks involved in everyday activities. Neuropsychological assessment is one method for diagnosing acquired disorders of neurocognitive function. Performance on neuropsychological assessments is related to critical life activities such as vocational success, community reintegration, and social autonomy. Neuropsychological measures are frequently part of the overall neurodiagnostic assessment, which includes other techniques such as neuroimaging. Neuropsychological exams are used routinely to assess the clinical relationship between neurologic functioning and cognitive/behavioral dysfunction, as well as to participate in differential diagnosis. The sensitivity of neuropsychological tests is such that they often
can identify patterns of impairment that are not determinable through other procedures, leading to appropriate treatment recommendations.

While neuropsychological assessments are used routinely to assess neurocognitive dysfunction, appropriate neuropsychological assessment services are often not readily accessible, due to a variety of factors that include long waiting lists, limited clinic hours, and remote geographical locations. Telemuropsychological approaches have been developed that allow the neuropsychologist to use video teleconferencing to administer neuropsychological measures remotely [1]. A limitation of this approach is that it requires a neuropsychologist to administer the tests and it misses out on valuable data (e.g., response time). Other solutions have involved the development of computer-automated neuropsychological assessments that administer visually mediated information (via a display monitor) without the need for an examiner [2]. However, the capacity for administering a core test battery has been limited by requiring patients to read instructions and by the need for tests that require verbal responses.

A potential resolution to these issues is the development of largely self-administering computerized neuropsychological assessments that can be administered remotely. Advances in virtual environments and virtual human technologies offer promise for overcoming these limitations [3-5]. A recently developed Avatar Administered Neuropsychological Testing (AVANT; see Figure 1) platform has been developed that includes a virtual reality environment in which a virtual human neuropsychologist administered neuropsychological assessment (including word list learning and recall measures).

**Figure 1:** Avatar Administered Neuropsychological Assessment Testing.

AVANT’s platform provides an interactive, primarily self-administering, experience, allowing for auditorily- and visually-mediated administration (via avatar-based verbal interaction) of cognitive assessment instructions to a real patient. Using speech recognition and a conversational avatar representing a virtual clinician, AVANT allows for avatar-administered neurocognitive assessments that augment a real clinician’s ability to assess patients. The AVANT’s speech recognition allows for virtual clinician-based administration of word-list learning, confrontation naming, and aural comprehension (see Figure 2). The current iteration of the AVANT word list learning and recall leverages recently validation approaches found in studies using virtual reality-based neuropsychological assessments of learning and memory [6-9].
The intent of the current study was to perform construct validation of a virtual human-administered neuropsychological assessment of learning and memory. This effort focused on validation of an automated language-based assessment system that is primarily self-administering with clearly presented directions using both visual illustrations and virtual human-based verbal instructions. The current study is an initial attempt at validating the Avatar Administered Neuropsychological Testing (AVANT) Memory Module, virtual human administered measures of learning and memory.

2. Methods

We acquired data on the implementation of a virtual human administered assessment of learning and memory (i.e., AVANT) in a normative sample that also received traditional paper-and-pencil tests. We aimed to assess the psychometric properties of the AVANT 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 aided our initial assessment of both the concurrent and divergent validity properties of this form of assessment.

2.1 Participants

The University’s Institutional Review Board approved the study. A total of 66 college-aged subjects participated in the study (see Table 1). The age range of participants was 19 to 27 years of age (age: M = 20.44; SD = 2.50). Participants were 64% female. The education range of participants was 12 to 15 years (education: M = 13.95; SD = 1.55).
Table 1: Demographics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Participants</td>
<td>66</td>
<td>100</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>Female</td>
<td>42</td>
<td>64</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Asian</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Hispanic</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>Caucasian</td>
<td>25</td>
<td>38</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>20.44</td>
<td>2.5</td>
</tr>
<tr>
<td>Education (years)</td>
<td>13.95</td>
<td>1.55</td>
</tr>
<tr>
<td>Estimated IQ</td>
<td>106.36</td>
<td>9.8</td>
</tr>
</tbody>
</table>

No significant differences were found for age, gender, or education. After informed consent was obtained, basic demographic information was recorded. Participants were also given a medical health history form to assess the presence of any mental or physical disorders that may have hindered their performance. No participants were excluded for responses given on this form.

2.2 Design and measures

To examine convergent and discriminant validity we compared the virtual human (i.e., AVANT) to face-to-face (i.e., human) administration of neuropsychological assessments. We hypothesized that the AVANT’s learning and memory scores would correlate with traditional neuropsychological measures involving learning and memory, but not with traditional measures involving potential confounds (i.e., executive functions; confrontation naming). Participants completed 1) paper-and-pencil neuropsychological tests administered under standard conditions (face-to-face with a human examiner); and 2) virtual human avatar administered neuropsychological tests of learning and memory.

Convergent validity analyses used the California Verbal Learning Test (CVLT) [10] as it is considered to have important learning and memory components and has been used clinically to estimate memory abilities. Discriminant validity measures included: (a) Color-Word Interference (Stroop) from the Delis–Kaplan Executive Function System (D-KEFS) [11]; and Picture Naming from the Repeatable Battery for the Assessment of Neuropsychological Status [12]. Participants completed both the AVANT tasks and paper-and-pencil neuropsychological tasks. Testing occurred in a quiet, climate-
controlled environment in a university-owned computer lab. The order in which the various tests were administered was counterbalanced across participants. Virtual reality learning and memory tasks. The AVANT Verbal Learning and Memory Test measures episodic verbal learning and memory. The measure examines encoding, recall and recognition of auditory-verbal information. Data is generated that reflects how much a patient learns, errors, and the strategies utilized. The AVANT Verbal Learning and Memory Test evaluates free and cued recall, serial position effects (including primacy and recency), semantic clustering, intrusions, interference and recognition. Two monitors were used: (a) one for displaying the Launcher application, which is used by the examiner (in another room) to monitor the virtual human’s administration of the list learning and recall tasks; and (b) another for displaying the participant’s view of the virtual environment. A scoring program was employed for data acquisition. This also allowed for key events in the environment to be logged and time stamped with millisecond temporal accuracy.

2.3 Data Analytics

All data were analyzed using SAS version 9.1. Descriptive statistics were calculated for participant demographics and for results of the AVANT and the criterion neuropsychological tests. Missing data were imputed by either mean substitution or last case carried forward. Correlations were computed among AVANT Verbal Learning and Memory Test and traditional neuropsychological measures assessing learning and memory. Two-tailed Pearson correlations were used.

2.4 Results

Convergent validity results indicated that AVANT word list learning correlated significantly with traditional neuropsychological learning (CVLT Trials 1–4; r = 0.47, p<0.001). AVANT memory correlated significantly with the traditional neuropsychological memory composite (CVLT free recall after a delay; r = 0.32, p<0.009; CVLT cued recall after a delay; r = 0.43, p<0.001).

Divergent validity results revealed no significant correlations between AVANT learning measures and executive function measures (DKEFS Inhibition/Switching (r = 0.12, p<0.33); or confrontational word retrieval (RBANS Picture Naming (r = 0.05, p<0.69). Likewise, no significant correlations between AVANT memory measures and executive function measures (DKEFS Inhibition/Switching (AVANT free recall r = 0.06, p<0.61; AVANT Cued Recall r = 0.09, p 0.44); and confrontational word retrieval (RBANS Picture Naming (AVANT free recall r = 0.01, p<0.91; AVANT Cued Recall r = 0.02, p<0.86).

3 Discussion

This study provides preliminary validation of the AVANT’s Verbal Learning and Memory. Convergent and discriminant validity were evaluated using neuropsychological tests chosen a priori. Results supported both convergent and discriminant validity. That is, findings suggest that the AVANT measures a capacity that is (a) consistent with that assessed by traditional paper-and-pencil measures involving learning and memory and
(b) inconsistent with that assessed by traditional paper-and-pencil measures assessing neurocognitive domains traditionally assumed to be other than learning and memory.

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 to ensure that the current findings are not an anomaly due to sample size. Additionally, the diagnostic utility of the AVANT’s Verbal Learning and Memory measures must be determined.

In summary, our goal was to conduct an initial pilot study to validate the AVANT’s Verbal Learning and Memory measures through comparison with standard neuropsychological tests for the assessment of healthy 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 are necessary to continue the process of test development and to fully establish the AVANT’s Verbal Learning and Memory measures as a contribution to existing assessment procedures for the diagnosis of memory decline. Although the AVANT’s Verbal Learning and Memory measures must be fully validated, current findings provide preliminary data regarding the validity of the VE as a memory measure.

References