Objective

Evaluation of performance validity is recognized as a necessary component of neuropsychological assessments recommended by both the American Academy of Clinical Neuropsychology (Helmkamper et al., 2009) and the National Academy of Neuropsychology (Bush et al., 2009). Without the use of performance validity tests (PVTs), accurate interpretation of test results obtained in a neuropsychological evaluation may be compromised. Traditionally, practitioners have relied on stand-alone measures of validity (e.g., Test of Memory Malingering [TOMM], Victoria Symptom Validity Test [VSVT], Validity Indicator Profile [VIP]), however there are several advantages to an embedded measure. They are efficient (i.e., they do not require additional time or testing) and can be applied retrospectively to previous testing (e.g., records review). Moreover, they allow for measurement of effort continuously throughout the battery (as effort may change during the course of the evaluation) and they are not recognizable as measures of effort (so they are less susceptible to coaching).

The Neuropsychological Assessment Battery (NAB) is an integrated neuropsychological battery used to assess cognitive skills in adults (Stern & White, 2003). A recent study developed two embedded PVTs within the NAB Scoring Module using the Attention and Executive Functioning modules (Race et al., 2021). Subtests that are most susceptible to poor effort (e.g., Digit Span and List Recognition) have been assessed throughout the battery (as effort may change during the course of the evaluation) and they are not recognizable as measures of effort (so they are less susceptible to coaching).

Participants

This study utilized archival data from 407 adult civil litigants referred for a neuropsychological evaluation at a private practice clinic in the western United States. Participants with noncredible performance (n = 47) were defined as those who failed two or more PVTs. Participants with credible performance (n = 259) were defined as those who did not fail any PVTs. Due to ambiguity associated with classifying performance for failing one PVT, those individuals (n = 101) were excluded, resulting in a final sample of 306 individuals. The final sample was 56% male (n = 171) with ages ranging from 18 to 85 years with a mean age of 42.7 years (SD = 15.8). Ethnicity and handedness were not specified for 63.4% (n = 194) and 63.7% (n = 195) of participants. Education for the total sample was reported as 7.5% (n = 23) for less than high school (HS) diploma, 26.5% (n = 81) for HS diploma (or equivalent), 28.8% (n = 82) for some college, and 26.8% (n = 82) for those with a bachelor’s degree or greater.

Procedure

Means and standard deviations were calculated for each subtest of the NAB Memory Module for both credible and noncredible groups. Independent-samples t-tests were conducted to examine mean differences on various subtests. Receiver-operating characteristic (ROC) curves were used to analyze the predicted accuracy of subtests of the NAB Memory Module. ROC curves were assessed by calculating the area under the curve (AUC), cutoff scores to maximize sensitivity and specificity were calculated.

Results

There were statistically significant differences in raw scores for the credible and noncredible groups on all 39 Memory Modules subtest scores. However, only 8 subtest scores demonstrated suitability for ROC analyses (AUC ≥ 0.80). List Learning A Immediate Recall, List Learning A Short Delayed Recall, List Learning A Long Delayed Recall, List Learning A Discriminability Index, Shape Learning Immediate Recognition, Daily Living Memory Immediate Recall, Daily Living Memory Delayed Recall, and Name/Address/Phone Delayed Recall. Means and standard deviations for the credible and noncredible groups as well as results from independent-samples t tests for each subtest are listed in Table 1.

Table 1. Results from Independent-Samples T Tests

<table>
<thead>
<tr>
<th>Raw score</th>
<th>Credible M (SD)</th>
<th>Noncredible M (SD)</th>
<th>t (df)</th>
<th>Sig</th>
<th>Cohen's d</th>
</tr>
</thead>
<tbody>
<tr>
<td>List Learning A Immediate Recall</td>
<td>22.6 (4.9)</td>
<td>15.7 (4.6)</td>
<td>-8.89 (300)</td>
<td>&lt; .001</td>
<td>1.45</td>
</tr>
<tr>
<td>List Learning A Short Delayed Recall</td>
<td>7.3 (2.5)</td>
<td>3.9 (2.1)</td>
<td>-4.86 (300)</td>
<td>&lt; .001</td>
<td>1.97</td>
</tr>
<tr>
<td>List Learning A Long Delayed Recall</td>
<td>7.1 (2.7)</td>
<td>3.5 (2.6)</td>
<td>-2.25 (300)</td>
<td>&lt; .001</td>
<td>1.36</td>
</tr>
<tr>
<td>List Learning A Discriminability Index</td>
<td>8.6 (2.8)</td>
<td>4.4 (3.5)</td>
<td>-7.50 (56.09)</td>
<td>&lt; .001</td>
<td>1.33</td>
</tr>
<tr>
<td>Shape Learning Immediate Recognition</td>
<td>17.8 (3.8)</td>
<td>12.4 (4.7)</td>
<td>-8.57 (300)</td>
<td>&lt; .001</td>
<td>1.26</td>
</tr>
<tr>
<td>Daily Living Memory Immediate Recall</td>
<td>42.4 (5.7)</td>
<td>32.8 (8.4)</td>
<td>-7.46 (52.64)</td>
<td>&lt; .001</td>
<td>1.34</td>
</tr>
<tr>
<td>Daily Living Memory Delayed Recall</td>
<td>13.8 (3.1)</td>
<td>8.6 (4.6)</td>
<td>-7.34 (52.48)</td>
<td>&lt; .001</td>
<td>1.33</td>
</tr>
</tbody>
</table>

For subtests where Levine’s test indicated unequal variances, the univariate F test was utilized.

Conclusions

These results provide preliminary psychometric evidence and clinical utility for use of various cutoff scores as a measure of embedded validity within the NAB Memory Module.

Limitations of the current study include small sample size and the potential of artificially inflated classification statistics due to spectrum bias resulting from exclusion of participants who failed one PVT (Schroeder et al., 2019). Spectrum bias occurs when the spectrum of clinical manifestations in the data classification statistics due to exclusion of participants who failed one PVT (Schroeder et al., 2019). Spectrum bias occurs when the spectrum of clinical manifestations in the data...
References


