Efficacy of the ANAM General Neuropsychological Screening battery (ANAM GNS) for Detecting Neurocognitive Impairment in a Mixed Clinical Sample

Jonathan Woodhouse, Daniel J. Heyanka, Jim Scott, Andrea Vincent, Tresa Roebuck-Spencer, Kristen Domboski-Davidson, Kerry O’Mahar, Russell Adams

University of Oklahoma Health Sciences Center, Department of Psychiatry & Behavioral Sciences, Oklahoma, OK, USA
Cognitive Science Research Center, University of Oklahoma, Oklahoma, OK, USA

demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Efficacy of the ANAM General Neuropsychological Screening battery (ANAM GNS) for Detecting Neurocognitive Impairment in a Mixed Clinical Sample

Jonathan Woodhouse1, Daniel J. Heyanka1, Jim Scott1, Andrea Vincent2, Tresa Roebuck-Spencer2, Kristen Domboski-Davidson1, Kerry O’Mahar1, and Russell Adams1
1University of Oklahoma Health Sciences Center, Department of Psychiatry & Behavioral Sciences, Oklahoma, OK, USA
2Cognitive Science Research Center, University of Oklahoma, Oklahoma, OK, USA

The Automated Neuropsychological Assessment Metrics (ANAM) is a computerized neuropsychological assessment battery that has demonstrated utility in a variety of clinical populations including multiple sclerosis, systemic lupus erythematosus, Parkinson’s disease, acquired brain injury, migraine headaches, and Alzheimer’s disease. This study utilized selected tests from the ANAM General Neuropsychological Screening Battery (ANAM GNS), a newly defined subset of tests from the broader ANAM library designed for general clinical assessment of cognition. ANAM GNS is an expansion of the ANAM Core battery which has been utilized in a military setting. The efficacy of the ANAM GNS was explored in a mixed clinical sample relative to well-established, traditional neuropsychological measure, the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). It was hypothesized that scores from the ANAM GNS would accurately predict participants as either impaired (n = 30) or normal (n = 113). Participants were grouped a priori based on RBANS Total Index scores with impairment defined as scores ≤ 15th percentile. Logistic regression analysis was conducted to evaluate the classification accuracy of the ANAM GNS. The predictor variables were the Throughput scores from seven selected ANAM GNS subtests. The full model significantly predicted impairment status, sensitivity was 81% and specificity was 89.1%. Overall classification rate was 87.9% and the Odds Ratio for the overall model was 34.65. Positive predictive value was 56.7% and negative predictive value was 96.4%. This study represents the first clinical data on the ANAM GNS, and documents that it has good concurrent and predictive validity with a well-established neuropsychological measure.

Keywords: Automated Neuropsychological Assessment Metrics; ANAM; Classification; Repeatable Battery for the Assessment of Neuropsychological Status; RBANS.

INTRODUCTION

In their landmark paper Kane and Kay (1992, pp. 1–2) summarized the many potential benefits of computer-based assessments. “The computer brings many advantages to testing. These include better standardization in administration and scoring, the ability to generate numerous alternate forms, precise stimulus control, the ability to track various subjects’ responses, increased cost efficiency in testing, and the ability to develop
large and accurate data bases.” However, one criticism of computer-based neuropsychological assessment is the frequent inadequacy of well-documented psychometric properties of these new technologies (Wild, Howieson, Webbe, Seelye, & Kaye, 2008).

Crook, Kay, and Larrabee (2009) recently reviewed several computerized assessment instruments with “the best psychometric foundations” (p. 84). The Automated Neuropsychological Assessment Metrics (ANAM) is one such computerized neuropsychological assessment battery that was developed by the Department of Defense (Reeves, Winter, Bleiberg, & Kane, 2007). Although the ANAM was originally developed for repeat assessment of military personnel, it has also demonstrated potential utility in a variety of clinical populations including multiple sclerosis, systemic lupus erythematosus, Parkinson’s disease, acquired brain injury, migraine headaches, and Alzheimer’s disease (Kane et al., 2007; Levinson, Reeves, Watson, & Harrison, 2005; Wild et al., 2008). The ANAM General Neuropsychological Screening Battery (ANAM GNS), a selection of specific clinically relevant tests from the larger ANAM library of tests, was utilized in this study. Although previous psychometric work has not been published on the specific ANAM GNS battery, it is comprised of the most frequently used ANAM tests in previously published studies. The ANAM GNS is also a direct expansion of the ANAM Core battery (previously referenced as the ANAM TBI battery) which has been utilized extensively in the Department of Defense (Vincent, Roebuck-Spencer, Gilliland, & Schlegel, 2012; Vincent, Roebuck-Spencer, Lopez et al., 2012). ANAM GNS begins with the full ANAM Core battery followed by additional tests assessing a broader range of cognitive domains.

While computerized neuropsychological screening is standard for sports concussion management (Maerlender et al., 2010), this new technology has also shown promise for the early detection of dementia (Wild et al., 2008), serial examinations of patients with epilepsy (Hoppe, Fliessbach, Schlegel, Elger, & Helmstaedter, 2009), and evaluating the cognitive effects of depression (Iverson, 2006). Kane and Kay (1992) noted that, “The objective for studies using ANAM in clinical settings has been to validate brief automated assessment techniques for screening and triage” (p. S115). Ongoing ANAM research has successfully correlated select ANAM subtests to traditional neuropsychological measures. For instance, a study by Kabat, Kane, Jefferson, and DiPino (2001) found moderate to strong correlations between the ANAM and traditional neuropsychological measures. The authors also provided a three-factor Principal Component Analysis solution (i.e., processing speed/efficiency, information retention, working memory) with ANAM subtests loading on the same factors as traditional measures of the same constructs. Likewise, Bleiberg, Kane, Reeves, Garmoe, and Halpern (2000) also demonstrated good construct validity between ANAM and traditional neuropsychological tests. ANAM has also been shown to be a good predictor of neuropsychological functioning in select clinical groups including lupus and multiple sclerosis (Roebuck-Spencer et al., 2006; Wilken et al., 2003).

The goal of this study is to advance the knowledge base of the ANAM GNS battery and to explore its efficacy for use as a cost-effective and time-saving screening measure in a clinical setting. The classification accuracy of ANAM GNS relative to a well-established traditional neuropsychological screening measure, the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) was evaluated in a mixed

1ANAM is distributed exclusively by Vista LifeSciences, Inc
It was hypothesized that the ANAM GNS would demonstrate concordance with the RBANS in predicting those individuals with cognitive impairment.

METHOD

Participants

Participants were recruited prospectively from patients referred to a teaching hospital for an outpatient clinical neuropsychological assessment; 147 participants, ages 18–85, volunteered to participate in this study. Minors and individuals with severe dementia and/or severe mental illness were excluded from the study. Before analyses were completed the dataset was explored for potential extreme outliers and to ensure that the statistical assumptions were met. Three participants were observed to have highly atypical scores (e.g., unusually low accuracy and response time scores that were highly discrepant from the rest of the sample). Closer examination of their data revealed that these individuals had severe cognitive impairment and exhibited a significant degree of confusion during computerized testing that compromised the validity of their data. These cases were thus excluded as outliers. An additional participant that reversed the buttons on the Mathematical Processing Test was identified. This participant’s responses were rescored and included in the final sample.

The remaining 143 participants were 91.6% White, 54.5% female, with a mean age and education of 50.59 (16.06) and 14.32 (2.76) years. On average, participants completed the ANAM GNS battery in 45 minutes (SD = 8.35) with a range of 30 to 67 minutes. Table 1 contains demographic information for the overall sample as well as the study groups (i.e., impaired and not impaired groups). Table 1 also contains the means and standard deviations for each group on the seven ANAM GNS subtests and the six RBANS indices. All patients were referred secondary to concerns or complaints about cognitive functioning. See Table 2 for a summary of post-evaluation neurologic and psychiatric diagnoses represented in this mixed clinical sample.

Procedure

This study received IRB approval and all participants provided informed consent. All participants were being seen for evaluation of their cognitive functioning, and thus received a comprehensive neuropsychological evaluation. If not already completed for clinical purposes, participants were administered the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). The RBANS was administered to assess core cognitive constructs commonly measured in a neuropsychological evaluation (i.e., immediate memory, delayed memory, attention, language, visuospatial skills) as well as global cognitive functioning. It was administered in the standardized format and utilized for diagnostic purposes while the ANAM GNS scores were strictly utilized for research purposes. The ANAM GNS was administered by a trained psychometrist in a standardized manner. The psychometrist remained in the room and provided clarification of instructions as needed. Data was stored in a secured electronic database.

Measures

Details regarding the traditional neuropsychological measures have already been described elsewhere (Lezak, Howieson, & Loring, 2004; Strauss, Sherman, & Spreen,
ANAM (Automated Neuropsychological Assessment Metrics); SRT (Simple Reaction Time), CDS (Code Substitution Learning), PRO (Procedural Reaction Time), MTH (Mathematical Processing), M2S (Matching to Sample), CDD (Code Substitution Delay), LRN (Logical Relations).

*ANAM Variables are presented as Throughput Scores. Impaired and non-impaired groups were statistically different for all cognitive tests.

**p < .001. †p < .01. ‡p < .05.

Table 2. Frequencies of neurologic and psychiatric disorders

<table>
<thead>
<tr>
<th>Etiological Group</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression/Anxiety</td>
<td>37%</td>
</tr>
<tr>
<td>Degenerative</td>
<td>9.8%</td>
</tr>
<tr>
<td>Pre-surgical</td>
<td>9.7% (includes healthy kidney donors)</td>
</tr>
<tr>
<td>Traumatic</td>
<td>5.6%</td>
</tr>
<tr>
<td>ADHD</td>
<td>4.5%</td>
</tr>
<tr>
<td>Neoplasm</td>
<td>3.5%</td>
</tr>
<tr>
<td>Vascular</td>
<td>2.1%</td>
</tr>
<tr>
<td>Inflammatory</td>
<td>2.1%</td>
</tr>
<tr>
<td>Congenital</td>
<td>2.1%</td>
</tr>
<tr>
<td>Toxic-Metabolic</td>
<td>1.4%</td>
</tr>
<tr>
<td>Seizure</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

ANAM can be conceptualized as a highly flexible library of neuropsychological tests. The ANAM GNS battery was designed to be a rapid computerized screen of major cognitive domains to assess for potential neurocognitive impairment. Bleiberg et al. (2000) and Kabat et al. (2001) demonstrated that ANAM appears to measure
similar cognitive constructs compared to traditional neuropsychological tests, most notably in the domains of cognitive efficiency, working memory, and resistance to interference. The ANAM GNS consists of the following 10 select measures: (1) Simple Reaction Time (motor speed), (2) Pursuit Tracking (psychomotor coordination), (3) Procedural Reaction Time (higher-order rapid responding), (4) Delayed Matching to Sample (short-term memory), (5) Simple Reaction Time Repeated (fatigue of motor speed), (6) Continuous Performance Test (attention and working memory), (7) Logical Relations (verbal based working memory), (8) Code Substitution Learning & Delayed Memory (immediate & delayed recognition memory), (9) Tower Puzzle (executive functioning), and (10) Mathematical Processing Test (working memory).

ANAM provides multiple variables for interpretation, including reaction time (measured in milliseconds), accuracy (measured by percentage correct), and Throughput. Throughput (TP) is a ratio of reaction time and accuracy which measures the number of correct responses per available unit of time. For the purposes of this analysis the TP variable was used given that it is robust and sensitive and is typically more normally distributed than the variables of reaction time and accuracy alone (Short, Cernich, Wilken, & Kane, 2007; Thorne, 2006). Distributions of the TP variables were examined for departures from normality. While some tests did show a moderate degree of skewness, the degree of departure was not of a magnitude to cause concern as Pearson correlations and the resulting t statistic as a test of $H_0: \rho = 0$ tends to be robust to non-normality (Edgell and Noon, 1984; Fowler, 1987; Havlicek and Peterson, 1977). Nonetheless we examined the correlations utilizing both parametric and non-parametric correlations and did not find any meaningful differences in the magnitude or significance of the resulting correlations. Therefore we have chosen to report statistics for untransformed data using parametric statistics.

Scores for the version of the Tower Puzzle subtests used for this study were not analyzed given that many participants had trouble understanding test directions and successfully completing this test. The version of ANAM GNS utilized in this sample was a beta version and it was anticipated that modifications would be made to improve the clinical applicability of the battery as needed. Based on the findings from this study, the Tower Puzzle has been revised given the observation that participants had difficulty understanding test instructions. The Pursuit Tracking and Continuous Performance Test were also not analyzed for this specific study because they do not generate Throughput scores, thus rendering comparison with the other subtests more difficult. In the final analysis, Simple Reaction Time, Procedural Reaction Time, Code Substitution Learning, Code Substitution Delayed, Delayed Matching to Sample, Logical Relations, and Mathematical Processing were investigated. Descriptions of each of these tests are provided in Table 3. Of note, these tests remain in an unaltered format in current versions of the ANAM GNS battery.

RESULTS

Participants were classified a priori as either impaired ($n = 30$) or not impaired ($n = 113$). Impairment was defined as RBANS Total Index scores $\leq 15$th percentile following the Heaton-NAB classification ranges (Stern and White, 2003). A Pearson product–moment correlation coefficient was calculated between age and the RBANS
Total. Since age was not significantly correlated with RBANS Total ($r = 0.125$, $p = 0.14$), it was deemed unnecessary to covary age in the subsequent analyses. Education was significantly different between the impaired and non-impaired groups, $t(141) = –2.47$, $p < .05$. A series of univariate ANCOVAs, adjusting for education, compared the impaired group to the non-impaired group on each of the RBANS and ANAM GNS variables. The impaired group performed significantly worse on all measures, typically at the $p < .001$ level. ANAM Code Substitution Learning was significant at the $p < .01$ level. ANAM GNS Simple Reaction Time and Code Substitution Delayed were significant at the $p < .05$ level (see Table 1). Partial correlations, adjusting for education, yielded moderately strong correlation coefficients between the selected ANAM GNS subtests and each of the RBANS Indices (see Table 4). A standard multiple linear regression demonstrated that a model of the seven previously described ANAM GNS TP variables significantly predicted the RBANS Total Index score, $F(7, 132) = 11.23$, $p < .001$, $R = .61$, $R^2 = .37$.

Classification accuracy of the ANAM GNS in distinguishing between impaired (RBANS Total $\leq$ 15th percentile) and non-impaired (RBANS Total $> 15$th percentile) participants was estimated through a logistic regression. The outcome variable in the model was impairment status. The predictor variables were the TP scores from the seven previously described ANAM GNS subtests. The full model significantly

<table>
<thead>
<tr>
<th>Test name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Reaction Time</td>
<td>Measures simple motor reaction time by having the user respond as quickly as possible to a target stimulus. Designed to assess and differentiate between pure reaction time and cognitive processing speed.</td>
</tr>
<tr>
<td>Code Substitution – Learning</td>
<td>Measures visual scanning, processing speed, attention, and learning by asking the user to compare a single symbol–digit pairing with a set of defined symbol–digit pairs presented at the top of the screen. The user is instructed to learn the symbol–digit pairing for a memory test to follow later in the battery.</td>
</tr>
<tr>
<td>Procedural Reaction Time</td>
<td>Measures attention and processing speed by having the user respond as quickly as possible to different sets of stimuli based on simple rules (e.g., press left mouse button if you see a 2 or 3 and right mouse button if you see a 4 or 5).</td>
</tr>
<tr>
<td>Mathematical Processing</td>
<td>Measures attention, basic computational skills, and working memory by asking the user to solve a single-digit arithmetic problem (e.g., “$5 – 2 + 3 =$”) involving two operations.</td>
</tr>
<tr>
<td>Matching to Sample</td>
<td>Measures visual spatial discrimination and working memory by presenting the user with a visual pattern for a specified period of time and then, following a brief delay, asking the user to select the previously seen pattern from two choices.</td>
</tr>
<tr>
<td>Code Substitution – Delayed Memory</td>
<td>Measures visual recognition memory by asking the user to compare a single displayed symbol–digit pair with the previously learned symbol–digit pairs presented earlier in the test battery (i.e., during the Code Substitution – Learning Test).</td>
</tr>
<tr>
<td>Logical Relations</td>
<td>Measures abstract reasoning and verbal syntax ability by asking the user to evaluate the truth of a statement (e.g., “&amp; comes after #”) describing the order of two symbols displayed on the display (e.g., “&amp; #”). The user presses designated buttons to indicate whether the statement is true or false.</td>
</tr>
</tbody>
</table>
predicated impairment status ($\chi^2 = 49.42$, df = 7, $p < .0001$). The model accounted for 46\% (Nagelkerke $R^2$) of the variance in impairment status. Sensitivity of the model predicting cognitive impairment was 81\% and specificity was 89.1\%. Positive predictive value was 56.7\% and negative predictive value was 96.4\%. Overall correct classification rate was 87.9\% and the Odds Ratio (OR) for the overall model was 34.65.

A correlation matrix (see Table 5) was composed to assess for multicollinearity amongst the ANAM GNS TP scores. As can be noted, there were statistically significant moderate intercorrelations between the seven ANAM GNS TP scores utilized in the aforementioned predictor model.

**DISCUSSION**

Consistent with previous research (see Bleiberg et al., 2000; Kabat et al., 2001), the ANAM GNS showed a strong relationship with traditional neuropsychological measures (i.e., RBANS Indices). As can be noted in Table 4, each of the seven ANAM GNS subtests was significantly correlated with the RBANS Total, Immediate Memory, Language, Attention, and Delayed Memory Indices. The RBANS Visuoconstructional Index was significantly correlated with Procedural Reaction Time,
Mathematical Processing, Matching to Sample, and Logical Relations. It was not correlated with Simple Reaction Time, Code Substitution Learning, or Code Substitution Delayed Memory. Interestingly, ANAM GNS variables demonstrated lower and less consistent correlations with the RBANS Visual/Construction and Language Indices, areas of cognition that are more focal in nature. Instead the correlations were stronger and more consistent between ANAM GNS variables and RBANS measures of Attention and Memory. When considered in combination in a multiple regression analysis, the ANAM GNS TP variables accounted for 37% of the variance in the RBANS Total Index. These results demonstrate the concurrent validity of ANAM GNS with traditional neuropsychological measures, with stronger concurrent validity for attention, memory, and overall cognitive functioning compared with more focal areas of cognition.

The primary hypothesis that the ANAM GNS will be able to significantly predict cognitive impairment and appropriately classify participants as either cognitively impaired or normal was supported. Performance on ANAM GNS correctly classified 87.9% of patients as impaired or not with a sensitivity of 81% and specificity of 89.1%. Likewise the overall model accounted for 46% of the variance in impairment status. Sensitivity might have been weakened to some extent by the possibility that impairment on the RBANS might have been driven by more focal language or visual-spatial impairments, cognitive domains that are measured less well by ANAM GNS. This is supported by stronger correlations between ANAM GNS and attention and memory measures than language and visual/construction measures. Additionally, a correlation matrix across the seven ANAM GNS TP scores revealed a moderate degree of multicollinearity, suggesting a less heterogeneous set of predictor variables. This is likely due to the fact that response speed is a large component of the TP score across all included ANAM tests. This issue should be examined more closely with a larger sample size. It is also possible that other subtests from the ANAM GNS that were not analyzed in this study may improve the classification accuracy.

There are several potential limitations of this study. A mixed clinical sample was used partly out of necessity because the sample sizes of each individual diagnostic group were too small to be analyzed independently. However, a mixed clinical sample does approximate more closely what is typically seen in a neuropsychological outpatient clinic. Even using the mixed clinical sample there were only 30 participants in the cognitively impaired group weighting this sample toward lower overall rates of cognitive impairment. It is possible that sensitivity would have been increased with a larger sample size and a greater range of cognitive impairment.

The definition of impairment, based on RBANS Total Score, has limitations. While the RBANS Total Score is considered to be a sensitive and reliable indicator of dementia, it may not necessarily reflect cognitive decline across etiological groups. Therefore it is certainly possible the participants were diagnosed as cognitively impaired based on their comprehensive neuropsychological assessment even if their RBANS Total Score was > 15th percentile.

It is noteworthy that one of the three excluded participants was diagnosed with frontotemporal dementia based on their comprehensive neuropsychological assessment. This raises an interesting question as to the potential limitations of computerized screening for some neuropsychological disorders in which clinical signs of the disorder
(i.e., disorganized thinking, disinhibition, variable attention, etc.) can interfere with the patient’s ability to complete the computerized assessment.

Future studies attempting to explore the classification accuracy of the ANAM GNS should use a larger sample size which would allow for examination of ANAM GNS’s ability to predict independent cognitive domains within a more comprehensive neuropsychological battery and to explore performance pattern differences on ANAM GNS in specifically defined diagnostic groups. Future studies should also consider evaluating the classification accuracy of specific subtests from the ANAM GNS, particularly those that were not included in this study. Finally, the differential applicability of computerized neuropsychological assessment for different neuropsychological disorders should be evaluated further.

In spite of these limitations, this study demonstrated that the ANAM GNS showed a strong relationship, good concurrent validity, and good diagnostic accuracy compared to a well-established traditional neuropsychological screening measure (RBANS). Further, as noted by Bauer et al. (2012), computerized neuropsychological assessment devices are potentially advantageous as a cost-effective and time saving screening tool. Nevertheless, there are multiple considerations one must take into account (e.g., purchase of a computer/assessment program/protocols, Current Procedural Terminology [CPT] codes) when evaluating the potential cost savings of a particular assessment measure. The ANAM GNS, with an overall correct classification rate of 87.9% in a mixed clinical sample demonstrated in this study, may very well be a useful measure in terms of clinical utility and potential cost-savings to clinicians.

REFERENCES


