ELSEVIER

Identifying and monitoring cognitive deficits in clinical populations using Automated Neuropsychological Assessment Metrics (ANAM) tests

Robert L. Kane [a,*], Tresa Roebuck-Spencer [b], Paul Short [c],
Michael Kabat [a,c], Jeffrey Wilken [d]

[a] *VA Maryland Healthcare System, United States*
[b] *National Rehabilitation Hospital, United States*
[c] *University of Maryland School of Medicine, United States*
[d] *Washington DC VA Medical Center, United States*

Abstract

In this article we review studies in which Automated Neuropsychological Assessment Metrics (ANAM) measures were used to screen for impairment in various clinical populations. These clinical groups include patients with multiple sclerosis, systemic lupus erythematosus, Parkinson's disease, Alzheimer's dementia, acquired brain injury, and migraine headache. Data are also presented from a group of outpatient referrals unselected with respect to clinical condition. Findings support the use of ANAM as a screening procedure for identifying the impaired patient.
© 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Computerized testing; Neuropsychology; Cognitive deficits

Although the Automated Neuropsychological Assessment Metrics (ANAM) system (Reeves et al., 1992) was developed from efforts within the DoD to monitor human performance changes in healthy individuals undergoing environmental challenges, it has increasingly been employed in the clinical arena. These clinical applications were motivated by the need to develop an instrument to screen patients at risk for neurocognitive impairment. It is neither practical nor cost-efficient to examine every patient at risk for neurocognitive deficits with a comprehensive neuropsychological battery. Hence, identifying patients in need of more detailed assessment is an important health care objective for patients and for the efficient use of health care resources. The objective for studies using ANAM in clinical settings has been to validate brief automated assessment techniques for screening and triage.

An additional motivation for developing ANAM for clinical use was that it was designed for repeated evaluations. Hence, it seemed well suited for the important task of monitoring changes in a patient's status and for assessing a patient's response to medications or other medical interventions.

In this article we review findings from investigations using ANAM with clinical populations. These studies all shared the common objective of validating ANAM as a clinical screening instrument and assessed the ability of relatively

---

* Corresponding author at: 127, VA Medical Center, 10 N. Greene Street, Baltimore, MD 21201, United States.
E-mail address: Robert.Kane@va.gov (R.L. Kane).

short ANAM batteries to identify impairment in at risk populations. In most cases, longer traditional batteries were used to separate impaired from non-impaired individuals. Patients were classified as impaired or not impaired based on their performance of these traditional measures; ANAM's sensitivity for detecting impairment was then judged against classifications based on traditional test performance. In other studies patients were compared against controls. In the case of migraine, patients were assessed pre, post, and during headache attacks. In these migraine studies ANAM was used to assess acute change and recovery. Specifically, in this paper we review findings from studies using ANAM to detect deficits in multiple sclerosis (MS), systemic lupus erythematosus (SLE), Parkinson's disease (PD), a mixed clinical sample, Alzheimer's disease (AD), acquired brain injury, and in patients suffering from migraine headaches.

1. Multiple sclerosis (MS)

MS is a demyelinating disorder affecting the central nervous system (CNS). Lesions can appear in various areas of the brain and brain atrophy is noted in some patients (Comi et al., 1993; Fulton et al., 1999; Gonzalez et al., 1994). Cognitive impairment can occur early in MS, and there is little relationship between the degree of physical and cognitive disability (Beatty, Goodkin, Hertsgaard, & Monson, 1990; Rao, Leo, Bernardin, & Unverzagt, 1991; van den Burg, van Zomeren, Minderhoud, Prange, & Meijer, 1987; Wishart & Sharpe, 1997). While the relationship between physical and cognitive disability is weak, the relationship is significant between cognitive deficits and quality of life.

Approximately 50% of patients with MS evidence cognitive changes. Early changes may be subtle and manifest as changes in processing efficiency rather than more substantial cognitive impairment. In some patients the extent of cognitive deficits may increase over time. Hence, there is a need in MS for methods of detecting the presence and progression of cognitive deficits that are both time efficient and cost effective.

Wilken et al. (2003) published initial findings from a multi-center study designed to assess ANAM as a tool for identifying neurocognitive deficits in patients with relapsing-remitting (RR) MS. All patients were given a battery of neurocognitive tests that included traditional neuropsychological measures frequently used with MS patients and selected ANAM measures. Participants diagnosed with MS were assessed prior to starting treatment with interferon-β1a and then at 6-month intervals for an additional 24 months. Published data reviewed here is for the initial assessment. Follow-up data will be published at the completion of the study.

The principal objective of this study was to assess the sensitivity of select ANAM measures for detecting neurocognitive deficits in MS. The design used traditional neuropsychological measures as the standard for assessing impairment. Patients' performance on ANAM was evaluated against their performance on these traditional measures. Statistical comparisons included correlation and regression analyses.

All patients were assigned an impairment index based on their performance of the standard neuropsychological measures including the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987); omission errors from the Kay Continuous Performance Test (unpublished); Digit Symbol and Block Design from the Wechsler Adult Intelligence Test-III (WAIS-III; Wechsler, 1997); Parts A and B of the Trail Making Tests (Reitan, 1992); total score from the Paced Auditory Serial Addition Test (PASAT; Gronwall, 1977); Wisconsin Card Sorting Test (WCST) perseverative responses (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); lexical (CFL; Benton, Hamsher, Varney, & Spreen, 1983) and semantic (Animal Naming; Goodglass & Kaplan, 1996) fluency scores; and dominant hand finger tapping speed (Reitan, 1979). The impairment index was computed in the traditional fashion as a ratio of the number of tests performed in an impaired range divided by the total number of included tests. A score was defined as impaired if it was greater than one standard deviation below the mean (Heaton, Grant, & Matthews, 1991). Based on prior neuropsychological research using impairment indices (Reitan, 1979), patients having an impairment index of 0.4 or higher were classified as impaired.

The ANAM measures used in this study included: Simple Reaction Time, Procedural Reaction Time, Code Substitution and Code Substitution Delayed Memory, Running Memory, Sternberg Memory Search, Logical Relations, Finger Tapping, and Mathematical Processing. In this study, the 4-item version of Mathematical Processing was used (see paper in this issue by Reeves et al.). ANAM tests were selected to assess cognitive deficits frequently reported in MS (Fischer, 2001).

The authors published tables of the correlation coefficients between traditional and select ANAM measures to assess similarities among the tests (Wilken et al., 2003). However, the focus of the study was to assess the ability of ANAM to identify individuals presenting as impaired on traditional measures and to discriminate these individuals from those performing normally. This was done statistically through two types of regression analysis.

Classification accuracy for ANAM in comparison to traditional measures was assessed using logistic regression. Using the impairment scores as described above, eight subjects were classified as impaired based on their performance of traditional neuropsychological instruments. ANAM measures identified seven of these as impaired. Forty subjects performed in the non-impaired range on traditional measures and 39 of these were classified as non-impaired by ANAM. ANAM's overall correct classification rate compared to traditional measures was 95.8%. This classification rate was obtained using three ANAM test scores: Mathematical Processing Accuracy, Code Substitution Delayed Memory Accuracy, and Running Memory throughput. When these three measures were placed into a step-wise regression, they accounted for 53% of the variance of the impairment index derived from the traditional test measures.

The authors (Wilken et al., 2003) also employed multiple regressions to determine the ANAM measures that best predicted the PASAT total score and the extent of shared variance with this measure. The PASAT has become one of the more widely used measures in MS research due to the demands it makes on processing speed and working memory. The ANAM parameters most highly correlated with PASAT performance were identified and entered into a stepwise multiple regressions with the PASAT total score as the dependent variable. At time one, 42% of the variance from the total score on the PASAT was accounted for by the combination of the Running Memory throughput and Mathematical Processing Accuracy scores.

2. Systemic lupus erythematosus (SLE)

Systemic lupus erythematosus is a chronic, often life-long autoimmune disease that ranges from mild to severe and primarily affects women. SLE involves multiple organ systems and most often manifests in the skin, joints, blood, and kidneys. The most common symptoms include joint pain, fatigue, and recurrent vascular injuries. Organ damage, often due to inflammation caused by autoantibodies and other disease processes tend to occur in a pattern of exacerbations interspersed by periods of remission.

Similar to MS, patients with SLE are at risk for cognitive changes but not all patients present with cognitive deficits. Reports of the incidence of central nervous system dysfunction in SLE have produced findings ranging from 50 to 90% (Ainiala, Loukkola, Peltola, Korpela, & Hietaharju, 2001; Futrell, Schultz, & Millikan, 1992; Johnson, DeLuca, Diamond, & Natelson, 1996; Kaell, Shetty, Lee, & Lockshin, 1986; McNicholl, Glynn, Mongey, Hutchinson, & Bresnihan, 1994). SLE-related CNS involvement can manifest in a wide spectrum of neurologic and psychiatric presentations, including stroke, seizures, peripheral neuropathy, chorea, dementia, psychosis, anxiety, and depression (American College of Reumatology, 1999). Cognitive deficits have been reported in SLE patients even without signs of overt structural brain changes (Carbotte, Denburg, & Denburg, 1995; Hanly et al., 1992; Hay et al., 1992) with between 15 and 38% of non-CNS SLE patients demonstrating cognitive impairment (Carbotte, Denburg, & Denburg, 1986; Hanly et al., 1992; Hay et al., 1992). Although results of neuropsychological studies do not indicate a clear cognitive profile in SLE patients, deficits in attention, immediate memory, psychomotor speed, visuospatial abilities, motor coordination, simple reaction time, and cognitive flexibility (Ainiala et al., 2001; Carbotte et al., 1995; Ginsburg et al., 1992; Hanly, Hong, Smith, & Fisk, 1999; Hay et al., 1994; Leritz, Brandt, Minor, Reis-Jensen, & Petri, 2000) have been reported. Hence, neurocognitive symptoms are reflective of the diverse ways in which SLE can present with subcortical features frequently noted. Because SLE-related cognitive changes are often subtle and fluctuating, repeatable and sensitive measures are needed to help detect the presence of cognitive impairment and clarify the course and determinants of cognitive dysfunction.

Holliday and co-workers employed tests from ANAM to track cognitive functioning in patients with SLE as part of the San Antonio Study of Lupus Neuropsychiatric Disease (SALUD), a 5-year longitudinal study of neuropsychiatric manifestations in SLE. All patients enrolled in this project were at least 18 years of age and met at least four of the 1982 revised criteria for SLE. Patients completed a standardized medical history; neurologic, rheumatologic and psychiatric examinations; and neuropsychological assessment. The neuropsychological assessment included ANAM measures along with traditional neurocognitive tests recommended by the American College of Rheumatology Consensus Group (1999). The sample of patients participating in this study were ethnically diverse, the majority of patients being Hispanic. Furthermore, there was a good representation of levels of disease status, educational attainment, and employment status. Brey et al. (2002) and Holliday et al. (2003) reported on 128 participants who received baseline evaluations as part of this project. Of the initial 128 patients, 67 were recruited to participate in neuropsychological testing.

Results of traditional neuropsychological testing revealed the expected pattern of impairment in SLE: slowed visuomotor speed, variable attention deficits, and retrieval deficits on memory testing (Brey et al., 2002; Holliday et al.,

2003). Performance on ANAM and traditional neuropsychological tests were compared using correlation and multiple regression analyses.

Because ANAM scores are not age/education corrected, individual ANAM throughput scores were correlated with raw scores on the traditional tests (Brey et al., 2002; Holliday et al., 2003). The overall pattern of correlations revealed that traditional tests of attention, memory, and psychomotor speed were correlated with ANAM measures. This finding is consistent with previous factor analytic studies (Bleiberg, Kane, Reeves, Garmoe, & Halpern, 2000). Specific correlations can be found in Table 2 of manuscript by Holliday et al. (2003).

Multiple regression analyses were conducted to determine the independent contribution of selected ANAM scores to the average *T*-score for all neuropsychological tests after controlling for age (Holliday et al., 2003). Together ANAM scores and age accounted for about 61% of the total variance in the overall *T*-score.

A key issue in using a brief automated battery to track patients is the ability of the brief battery to identify individuals demonstrating cognitive impairment. Roebuck-Spencer et al. (2004) used logistic regression (LR) to assess the ability of ANAM measures to classify patients based on their performances on standard neuropsychological measures. Raw scores from the traditional measures were converted to *z*-scores based on published norms and averaged in order to calculate a global level of functioning. For the most part SLE patients in this sample obtained test scores that were well within normal limits (mean global *z*-score = -0.24) with approximately 25% of individuals performing in the low average or mildly impaired range.

For their logistic regression analysis, Roebuck-Spencer et al. (2006) used throughput scores for all ANAM measures with the exception of Simple Reaction Time. For this latter measure they used the mean response time score. After controlling for premorbid ability level, ANAM measures classified 80% of individuals as either having no or probable cognitive impairment as determined by traditional neuropsychological measures, with 76.2% sensitivity and 82.8% specificity. The model was statistically significant: $\chi^2(9) = 23.45$, $p < 0.01$. Importantly, the ability of ANAM to predict neuropsychological functioning remained even when the effects of self reported sleepiness and depression were controlled.

It is important to take into account cultural issues when screening for neurocognitive impairment. In the SALUD study, Hispanic patients had lower scores than non-Hispanic patients on the cognitive domains of attention/concentration and verbal functions measured by traditional tests. Hispanic ethnicity was not related to ANAM summary measures (Holliday et al., 2003). None of the individual ANAM scores were significantly related to Hispanic ethnicity, although there were trends for Hispanics to perform more poorly on Code Substitution Immediate Memory Throughput ($p = 0.055$), Accuracy ($p = 0.075$) and Reaction Time ($p = 0.075$) scores. Similar findings were seen for English language skills assessed by the Woodcock–Munoz Language Survey. Whereas English competency was related to traditional neuropsychological measures, it was not correlated with any ANAM measures. Furthermore, none of the ANAM summary measures was significantly correlated with education, and of the individual test scores, only Code Substitution Immediate Memory Accuracy and Variability scores were correlated with education. The vast majority of ANAM measures were significantly correlated with age, with younger patients performing better than older patients as expected.

In sum, ANAM measures obtained in a group of SLE patients correlated significantly with scores from the traditional neuropsychological test battery recommended by the American College of Rheumatology Consensus Group. Further ANAM was a good predictor of overall cognitive status in patients with SLE as assessed by traditional neuropsychological testing. More importantly, data with an ethnically diverse SLE sample indicated that ANAM was less sensitive to the confounding effects of education, English language proficiency, and ethnic differences than the traditional battery. Finally, it appeared that ANAM's sensitivity to cognitive change should make it a useful tool for identifying and tracking cognitive deficits in SLE patients.

3. Parkinson's disease (PD)

PD has been associated with various cognitive changes and with a number patients developing dementia as the disease becomes more chronic (Aarsland, Tandberg, Larsen, & Cummings, 1996). There are consistent pathological changes in PD coupled with findings that vary from patient to patient. Common to all patients with PD is a moderately severe nerve cell loss ($\geq 60\%$) in the pars compacta of the substantia nigra (Gibb, 1992). Lewy bodies, evidencing neural degeneration, can be found is some remaining cells (Soukup & Adams, 1996). Lewy bodies have also been found in other brain areas outside of the substantia nigra. These areas include: the locus coeruleus, ventral tegmental area,

Table 1
Traditional neurocognitive tests indices included in the impairment index

Measure	Functional domain
WAIS-III Working Memory Index	Working memory
WAIS-III Processing Speed Index	Processing speed
WAIS-III Block Design	Design analysis and synthesis
WAIS-III Picture Completion	Visual scanning and discrimination
WAIS-III similarities	Abstract verbal reasoning
California Verbal Learning Test-2nd Edition Total Score	Verbal acquisition
California Verbal Learning Test-2nd Edition Long Delay Free Recall	Verbal recall
California Verbal Learning Test-2nd Edition Discrimination	Verbal recognition
Rey Complex Figure Delayed Recall	Non-verbal recall
Animal Naming	Categorical fluency
Boston Naming Test	Confrontation naming
Wisconsin Card Sorting Test Conceptual Response	Problems solving
Wisconsin Card Sorting Test Perseverative Response	Executive function
Stroop Word Color Test Interference	Executive function
Trails A total time	Psychomotor processing speed
Trails B total time	Executive function

nucleus basalis of Meynert, raphe neuclei, thalamus, cerebral cortex, and throughout the autonomic nervous system (Soukup & Adams, 1996). Studies have shown that significant neuronal pathology occurs prior to the manifestation of clinical symptoms in PD. It has been estimated that there is a 70–90% reduction in striatal dopamine and a 60–70% loss of neurons in the substantia nigra prior to the manifestation of symptoms (Koller & Montgomery, 1997). Hence, it is not clear at what point cognitive changes may be detected in this disease.

There are a several investigations underway exploring the use of ANAM in PD. Preliminary findings from a study being conducted at the University of Maryland Medical Center were recently reported by Short and Kane (unpublished report) and are summarized here. The specific focus of this study was on cognitive changes in PD patients whose symptoms were rated as mild and who did not present with dementia as assessed with the Mini Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975). The research question was can ANAM be used to efficiently detect cognitive impairment in patients at risk secondary to PD. To address this question PD patients meeting criteria were assessed with ANAM and with a comprehensive neuropsychological test battery. Performance on the traditional neuropsychological test battery was used to divide the PD patients into two groups: those with cognitive impairment and those without. Spouses of PD patients were also tested with ANAM and served as non-PD controls. At the time of the initial analysis, 43 patients completed both ANAM and the traditional neuropsychological examination and 39 controls had completed ANAM.

Since patients with mild PD may or may not present with cognitive changes, performance on the traditional neuropsychological measures was used to classify patients as impaired or non-impaired. The traditional metrics used to assess PD patients in this study are listed in Table 1.

An impairment index approach was used to place patients into impaired or non-impaired categories based on these traditional test measures. Performance on each administered neuropsychological measure was incorporated into a comprehensive impairment index in the following manner. Any score poorer than one standard deviation below the mean received a score of one with better performances scored as zero. Tests for which all subjects performed within normal limits (e.g. Controlled Oral Word Test) were excluded from further analysis. Based on probability models by Ingraham and Aiken (1996), a score of four was used as a cut-point for group membership. This number was chosen because the likelihood for obtaining at least 4 of 16 scores one standard deviation below the mean was less than 20%. Thus, 12 participants were designated as impaired (35% of sample), with indexes ranging from 4 to 11. PD patients obtaining an impairment index score of 3 or below were designated PD normal. Controls and PD normal individuals were pooled for some analyses. The final groups for comparison consisted of 22 PD normal (NPD) patients, 12 impaired PD patients (IPD), and 36 normal controls (NC). The groups differed in age: IPD = 68.8 (7.6), NPD = 58.8 (9.7), NC = 62.1 (11.8). The IPD group was significantly older than the other two groups.

Table 2
Summary of regression models predicting throughput

ANAM	F-test	p	Adjusted R^2	PD group	p	Age	p
CDD	$F(3,63) = 6.29$	0.0008	0.19	$t = -2.18$	0.033	$t = -2.96$	0.004
CDS	$F(3,66) = 15.63$	<0.0001	0.39	$t = -4.47$	<0.001	$t = 10.63$	<0.0001
CH2	$F(3,61) = 16.02$	<0.0001	0.41	$t = 10.56$	<0.0001	$t = -2.44$	0.018
CPT	$F(3,63) = 12.73$	<0.0001	0.35	$t = -3.38$	0.0013	$t = -3.95$	0.0002
LRS	$F(3,66) = 6.62$	0.0005	0.20	$t = -2.14$	0.036	$t = -2.85$	0.006
MSP	$F(3,66) = 14.17$	<0.0001	0.36	$t = -2.31$	0.024	$t = -4.80$	<0.0001
MTH	$F(3,64) = 10.75$	0.0036	0.15	$t = -1.63$	0.110	$t = -2.87$	0.006
PRO	$F(3,64) = 10.75$	<0.0001	0.30	$t = -3.76$	0.0004	$t = -2.73$	0.008
STN	$F(3,64) = 7.31$	0.0003	0.22	$t = -2.46$	0.017	$t = -2.52$	0.014

CDD: Code Substitution Delayed Memory; CDS: Code Substitution; CH2: Two Choice Reaction Time; CPT: Running Memory; LRS: Logical Reasoning; MSP: Delayed Matching to Sample; MTH: Mathematical Processing; PRO: Procedural Reaction Time; STN: Sternberg.

Table 2 presents results for a series of regression analyses with group [(IPD versus (NPD & NC)] and age predicting throughput scores for specific ANAM tests.

Adjusted R^2 values indicated that age and group membership explained between 15 and 41% of the variance in throughput scores. Age was a significant predictor for all models, whereas membership in the impaired PD group was a significant predictor for all but the Mathematical Processing task. Although the data do not appear in this table, membership in the unimpaired PD group was not a significant predictor for ANAM throughput scores, as each of the p-values for these t-tests exceeded 0.75.

As noted in the introduction of this paper, an anticipated use of brief screening batteries in clinical settings is to identify which at risk patients are demonstrating evidence of neurocognitive impairments. In this context it would be useful to implement a summary score that captured performance on the ANAM battery for identifying patients evidencing impairment. To this end, Short and Kane developed a weighted throughput score to serve as a single metric capturing performance across an entire ANAM battery. They then tested this weighted score in this sample of PD patients. Since ANAM tasks differ, typical throughput scores can vary by task. The weighting procedure permitted the throughput score for each test to be counted equally rather than having a specific test contributing disproportionately to the final score. Short and Kane termed this weighed throughput score an Index of Cognitive Efficiency (ICE) score. Higher scores are indicative of better performance.

Mean ICE values for each of the study groups appear in Table 3. The differences between groups were tested with a regression model predicting ICE scores and were significant, $F(2,57) = 13.05$, $p < 0.0001$. Moreover, while membership in the impaired group produced significantly lower scores, $t = -4.66$, $p = 0.0001$, the PD normal group was similar to the control group, $t = 0.58$, ns. However, as indicated earlier, the groups differed in mean age, with the NPD group somewhat younger than the IPD group. Controlling for age, group membership continued to significantly influence ICE performance, $F(2,57) = 10.25$, $p = 0.0001$. The PD impaired group continued to perform much more poorly, $t = -4.38$, $p = 0.005$. When age was accounted for, group membership explained about 34% of the ICE score variance. As the boxplots in Fig. 1 suggest, the PD normal and control groups appeared to be virtually indistinguishable, displaying similar levels of median ICE performance.

Thus, it appeared that the ICE score may provide a reasonable screening method for detecting broad cognitive change in PD patients. As additional data are acquired for PD patients exhibiting impairment on traditional neurocognitive testing, it may become possible to establish cut scores that would allow clinicians to screen PD patients who might require more extensive neurocognitive evaluation.

Table 3
Weighted index of cognitive efficiency by study group

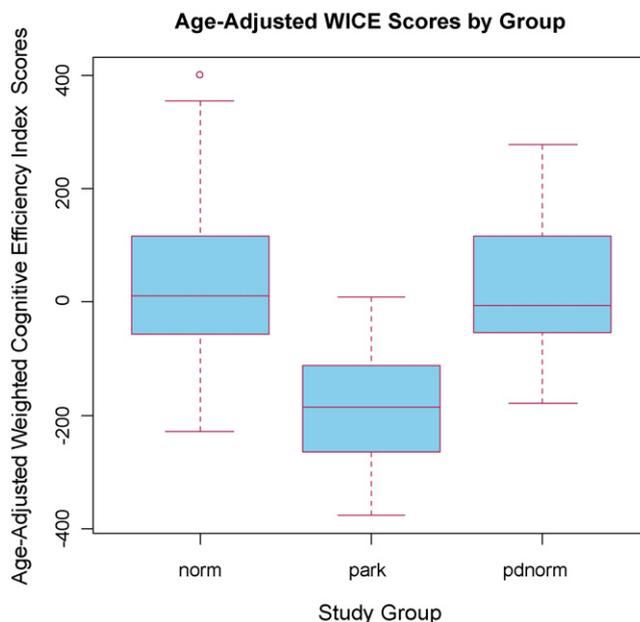|	Mean	S.D.
Control	878.97	161.29
PD normal	906.81	177.08
PD impaired	598.17	156.25

PD: Parkinson's disease patients.

Fig. 1. ICE performance on scores adjusted for age effects.

4. Alzheimer's dementia (AD)

According to the Alzheimer's Disease Education and Referral Center, approximately 4.5 million Americans suffer from AD with the frequency of occurrence increasing with advancing age. While a number of mental status examinations are employed in clinics, a brief screen designed to be repeatable and to track changes would be useful in the early detection and tracking of this disorder.

Levinson, Reeves, Watson, and Harrison (2005) reported findings from a pilot study using ANAM tests to assess cognitive impairments associated with mild AD. These authors compared the performance of eight individuals with AD to that of eight geriatric controls. There was no significant difference in age between the two groups: the AD patients had a mean age of 77.9 (S.D. = 5.17); the controls had a mean age of 75.6 (S.D. = 5.18). The MMSE scores for the controls (mean = 29.6, S.D. = 0.8) were within the normal range, while the MMSE scores for the AD group were consistent with mild dementia (mean = 24.5, S.D. = 3.55). The groups differed with respect to their scores on the Geriatric Depression Scale (GDS), although neither group appeared depressed on this instrument.

The ANAM battery selected for this study was comprised of six tests assessing response time, spatial perception, and working memory: Simple Reaction Time, 2-Choice Reaction Time, Delayed Matching to Sample, Running Memory, Simultaneous Spatial Processing, and Sternberg. Univariate tests using accuracy scores indicated that the patients were significantly impaired on three of the six tests: Delayed Matching to Sample, Running Memory, and Sternberg; all Fs $(1,14) > 11.00$; all p values < 0.01. Measures showing statistical significance (Delayed Matching to Sample, Running Memory, Sternberg) appeared to have a working memory component (Levinson et al., 2005). A discriminant function analysis using accuracy scores produced a 93.8% correct classification rate with one AD patient incorrectly classified as a control.

Using throughput rather than accuracy produced a greater number of scores that differed significantly between the two groups and yielded a higher rate of correct classification. Univariate analysis indicated that patients performed more poorly on all tests with the exception of Simple Reaction Time; Fs $(1,14) > 4.75$, p values < 0.05. A disciminant function correctly classified 100% of the participants (Levinson et al., 2005).

As noted, the ANAM battery employed in this study was composed of six tests. This favorable variable to subject ratio may well have contributed to the discriminatory power of the battery. Consequently, the authors followed their initial analysis with a logistic regression employing just two tests from the ANAM battery. Using the throughput scores

from Sternberg Memory and Running Memory, they were able to identify all eight patients as impaired and all controls as normal.

This study by Levinson and co-workers was a pilot study that involved 16 participants. Nevertheless, there are two aspects of the study that are encouraging with respect to using ANAM as a brief screen for dementia. First, the patients used in the study appeared to be in the earlier stages of their disease based on their MMSE scores. Second, ANAM was highly accurate in discriminating patients from controls using only two tests from the battery. Replication using a larger group of patients and non-impaired controls is needed.

5. General clinical population

The Spaceflight Cognitive Assessment Tool for Windows (WinSCAT; Kane, Short, Sipes, & Flynn, 2005) is a special subset of ANAM configured for NASA. This application is similar to others reviewed in this paper in that the purpose of WinSCAT is to screen for and monitor potential changes in neurocognitive status. In order to enhance information obtained from routine assessments, Kane and co-workers incorporated WinSCAT into the assessments of a group of patients referred for clinical examination. Later, they did a retrospective review of the data set treating WinSCAT as a screening instrument that could be used to identify impairment and the traditional measures as the standard for validating the screen. This analysis has not been published but is reviewed here.

Ninety patients completed WinSCAT as part of their evaluation but the referral question determined the selection of the traditional test measures. Hence, the choice of traditional neurocognitive instruments was based on clinical discretion and the precise battery varied from patient to patient. A sub-sample ($n = 31$) received a relatively standard battery of neurocognitive tests allowing for statistical comparisons between a consistent group of traditional neuropsychological measures and this specific group of ANAM tests. This select sample was demographically similar to the larger group: male (83%), mean age = 46.4, S.D. = 8.3, and mean education level = 13.4, S.D. = 2.6 years. Ethnicity was also similar to the larger group: African-American (56.3%); Caucasian (33%); and Hispanic (8.3%).

WinSCAT is comprised of the following ANAM tests: Code Substitution, Code Substitution Delayed Memory, Running Memory, Mathematical Processing, and Delayed Matching to Sample. Throughput scores were used in their raw units. To facilitate interpretation reaction time raw scores were converted to 100ths of a second. Accuracy raw scores were treated as percentages.

A full traditional battery was operationally defined as evaluating five cognitive domains: attention, working memory, learning, recall, and executive functioning. Each domain was assessed by a consistent set of tasks and there were multiple measures of each putative domain. Attention was assessed with the Trail Making Test Part A *T*-score and WAIS-III Digit Symbol scaled score. Working memory was assessed from WAIS-III scaled scores comprising the Working Memory Index triad: Arithmetic, Letter–Number Sequencing, and Digit Span. The learning domain was evaluated with *T*-scores from the CVLT total recall score across five trials and learning scores from the Heaton Story and Figure Memory Tests (Heaton et al., 1991). Recall was assessed using delayed recall on the same three tasks. For the CVLT, we used the 20-min delay interval. Finally, executive function was evaluated with *T*-scores for the Trail Making Test part B and the conceptual level responses score from the WCST. Scores for each test were recoded in standard deviations (S.D.) above and below the expected mean score. Scores less than one S.D. below the mean were coded 0. Scores between one and two S.D. below the mean were coded 1, and scores 2 or more S.D. below the mean were coded 2. These groupings roughly corresponded to non-impaired, marginally to mildly impaired, and impaired for a given test. In all cases, the original metric underlying the impairment index was age-corrected.

In order to look at the sensitivity of the ANAM tests incorporated into WinSCAT as screening measures in a general clinical sample, a composite global impairment index was calculated by summing impairment rating scores for the 13 conventional tests. The subsequent scores ranged from 0 to 16 and were relatively normal in their distribution (mean 7.25 ± 3.94, median 7). Since the entire sample was composed of clinical referrals, patients were stratified according to level of impairment. In order to determine whether ANAM tests in the battery were sensitive to impairment severity, the sample was divided at the median into two groups of approximately equal sizes, with the low-impairment group operationally defined as having impairment scores below 7. A multivariate analysis of variance (MANOVA) was performed to evaluate systematic differences in throughput scores by group. An exploratory discriminant analysis was then performed to determine how well the five throughput scores performed in classifying individuals into their defined groups (Stevens, 1996).

Table 4
Post hoc univariate analysis of WinSCAT throughput scores

Test	T-score	Low impairment, mean (S.D.)	High impairment mean (S.D.)	$F(1,29)$	Significance
CDS	Throughput	43.56 (8.29)	33.20 (4.16)	18.95	<0.001
CDD	Throughput	31.31 (12.05)	26.33 (6.82)	1.97	0.17
CPT	Throughput	84.50 (16.17)	58.87 (19.03)	16.40	<0.001
MTH	Throughput	19.88 (5.73)	17.27 (5.82)	1.58	0.22
MSP	Throughput	25.63 (8.71)	18.47 (7.03)	6.29	0.02

CDS: Code Substitution; CDD: Code Substitution Delayed Memory; CPT: Running Memory; MTH: Mathematical Processing; MSP: Delayed Matching to Sample.

5.1. Multiple regression analysis

Throughput scores from the five ANAM tests were entered into a regression model predicting composite global impairment index scores. The results of this analysis were significant, $F(5,25) = 3.37$, $p = 0.02$. The model predicted 40% of the variance in impairment scores although no single test was a significant predictor. Thus, the battery as a whole was sensitive to impairment level in a mixed clinical population with a variety of presenting problems.

5.2. Multivariate analysis

The MANOVA examining group differences (low versus high impairment) on the composite of ANAM throughput scores was significant, Hotelling's $T = 1.02$, $F(5,25) = 5.10$, $p = 0.002$. This finding appears fairly robust as power to detect true differences is greatly attenuated when there are fewer than 20 subjects per group in a multivariate investigation (Stevens, 1996). Thus ANAM measures employed in WinSCAT were sensitive to severity of impairment.

As is evident from the analysis presented in Table 4, increased level of impairment influenced Code Substitution, Running Memory, and Delayed Matching to Sample. Although Mathematical Processing and Code Substitution Memory exhibited similar trends, the power given the sample size to detect a true group difference in this model was low (0.22 and 0.27, respectively). For all ANAM tests in the WinSCAT battery, throughput scores were greater (better) for the less impaired group.

5.3. Discriminant function analysis

The initial discriminant function analysis, for which the sample was split at the mean, produced a significant distinction between groups, Wilk's lambda = 0.50, $\chi^2(5) = 18.65$, $p = 0.002$. The single function produced an eigenvalue of 1.02. An evaluation of the structure matrix suggested that Code Substitution Learning (0.80), Running Memory (0.74), and Delayed Matching to Sample (0.46) throughput scores were most highly correlated with the subsequent function. The remaining two tasks were only moderately correlated.

As is evident from Table 5, the derived canonical correlation exhibited good screening properties. Moreover, results from this sample presenting with a variety of clinical concerns parallels findings from other studies reviewed in this paper and elsewhere in this issue. ANAM tests employed in WinSCAT and in other ANAM batteries appear to provide good sensitivity and specificity across studies and in different clinical populations.

Table 5
Group classification based on discriminant function derived from WinSCAT tests

Impairment index based on traditional tests	Predicted WinSCAT impairment classification		
	Low	High	Percentage correct
Low	13 subjects	3 subject	81.3
High	2 subject	13 subjects	86.7
Overall			83.9

6. Acquired brain injury (ABI)

With ABI, comprehensive examination is typically indicated more than time efficient screening. This is especially the case when questions of rehabilitation planning are involved. However, data where ANAM was used with head injured individuals provides additional data regarding the battery's sensitivity to the presence and degree of impairment.

Levinson and Reeves (1997) studied 22 individuals enrolled in a traumatic brain injury rehabilitation program. Participants were classified as marginally impaired, mildly impaired, or moderately impaired, based on their performances of other neurocognitive assessment measures and staff ratings. These individuals were not simply concussed, but required rehabilitative services because of their injuries. Participants ranged between 18 and 64 years of age with a mean age of 38. Most participants suffered traumatic brain injury secondary to motor vehicle accidents. However, two suffered vascular accidents. Consistent with previous studies, throughput scores were superior to accuracy scores in distinguishing between individuals suffering from greater and lesser degrees of impairment.

Participants in the study were assessed on two occasions separated by a 2–3-month interval. Differences among groups were more pronounced at time-one. At time-one, it was possible to classify participants into groups using ANAM throughput scores with 91% accuracy. Combining mild and moderately impaired patients into one group and placing marginally impaired patients into a second group produced a time one classification rate of 100%. At time two, it was possible to classify individuals into one of three impairment groups with 86.36% accuracy using ANAM throughput scores. Results of this pilot study indicated that ANAM throughput scores were sensitive to levels of impairment in this group going through a rehabilitation program.

7. Migraine headache

ANAM findings in patients with migraine headaches are reviewed in two papers contained in this supplement: one related to studies of cognitive effects and side effects of pharmaceuticals and the second exploring the use of change indexes to monitor alterations in cognition. However, these headache studies are briefly referenced here since they demonstrate the clinical application of using a repeatable test battery to monitor changes in patients' conditions.

Farmer, Cady, Bleiberg, and Reeves (2000) reported test findings from 10 migraineurs who were studied at baseline, during the occurrence of a migraine attack, and during recovery following pharmacological intervention. These investigators used a short battery consisting of Simple Reaction Time, Delayed Matching to Sample, Mathematical Processing, and Running Memory. Throughput scores for all four tests included in the battery were significantly lower during a migraine event than at baseline, with improvement in test scores seen with recovery. Farmer et al. (2001) reported similar findings from a study of 28 women mirgaineurs using the same ANAM tests noted in the previous study.

This implementation of ANAM required establishing baseline values, monitoring patients during their migraines, and then re-assessment once the attack subsided. This type of clinical application could only be accomplished with a test battery that was brief and designed for multiple test administrations. Hence, these studies by Farmer and her colleagues provide a model for clinical monitoring of patients experiencing acute changes in their conditions either as a result of the course of their disorders or secondary to a medical intervention.

8. Conclusions

In this article we summarized findings from studies where selected ANAM measures were used to assess various patient populations. The primary focus of these studies was the use of a brief automated battery to identify impairment in groups at risk for alterations in cognition. While lengthier batteries composed of traditional test measures have proven useful in characterizing patient deficits, they are not practical when screening for deficits in at risk patient populations nor were they designed for repeated administrations.

In all studies, ANAM measures were sensitive to cognitive changes associated with neurological disorders. Current data do not suggest that brief ANAM batteries provide the same information as more extensive neurocognitive assessments. However, data are consistent in suggesting that brief ANAM batteries are both sensitive and specific in identifying patients with neurocognitive difficulties.

Within the field of neuropsychology, there is a need for time efficient computerized batteries capable of identifying patients who are experiencing neurocognitive problems. In some instances, this brief assessment may be sufficient to

meet the clinical need. In other cases brief automated procedures may identify individuals requiring more extensive examination. Uses of brief automated examinations include: screening and triage; assessing conditions where there are notable deficits in attention, working memory, and processing speed; monitoring disease progression; and assessing the effects of treatment. Primarily, this paper presented data related to the first issue, that of screening patients for indications of neurocognitive impairment. Data presented in other papers in this issue address in more detail the use of ANAM to monitor change and to assess the effects of medications. Considering data from this and companion articles in this issue, several conclusions present themselves. First, it appears that a brief group of ANAM measures can effectively screen for cognitive changes associated with neurological and medical disorders. Second, ANAM seems to be an effective and efficient tool for monitoring changes in patients' conditions. When ANAM was first brought into the clinical arena, the expectation was that a battery designed to assess performance efficiency and to be repeatable would have a number of important clinical applications. The data reviewed in this and in other papers in this supplement appear to support this contention. Additional work including larger scale studies are needed to further develop ANAM as a clinical tool. Work completed to date argues for the importance of these follow-up studies and for ANAM's potential as a clinical instrument.

References

(1999). The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. *Arthritis and Rheumatism*, *42*(4), 599–608.

Aarsland, D., Tandberg, E., Larsen, J. P., & Cummings, J. L. (1996). Frequency of dementia in Parkinson disease. *Archives of Neurology*, *53*(6), 538–542.

Ainiala, H., Loukkola, J., Peltola, J., Korpela, M., & Hietaharju, A. (2001). The prevalence of neuropsychiatric syndromes in systemic lupus erythematosus. *Neurology*, *57*(3), 496–500.

Beatty, W. W., Goodkin, D., Hertsgaard, D., & Monson, N. (1990). Clinical and demographic predictors of cognitive performance in multiple sclerosis: Do diagnostic type, disease duration, and disability matter? *Archives of Neurology*, *47*, 305–308.

Benton, A. L., Hamsher, K., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. Iowa City: AJA Associates.

Bleiberg, J., Kane, R. L., Reeves, D. L., Garmoe, W. S., & Halpern, E. (2000). Factor analysis of computerized and traditional tests used in mild brain injury research. *The Clinical Neuropsychologist*, *14*(3), 287–294.

Brey, R. L., Holliday, S. L., Saklad, A. R., Navarrete, M. G., Hermosillo-Romo, D., Stallworth, C. L., et al. (2002). Neuropsychiatric syndromes in lupus: Prevalence using standardized definitions. *Neurology*, *58*(8), 1214–1220.

Carbotte, R. M., Denburg, S. D., & Denburg, J. A. (1986). Prevalence of cognitive impairment in systemic lupus erythematosus. *The Journal of Nervous and Mental Disease*, *174*(6), 357–364.

Carbotte, R. M., Denburg, S. D., & Denburg, J. A. (1995). Cognitive dysfunction in systemic lupus erythematosus is independent of active disease. *Journal of Rheumatology*, *22*(5), 863–867.

Comi, G., Filippi, M., Martinelli, V., Sirabian, G., Visciani, A., Campi, A., et al. (1993). Brain magnetic resonance imaging correlates of cognitive impairment in multiple sclerosis. *Journal of the Neurological Sciences*, *115*(Suppl.), S66–S73.

Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *The California Verbal Learning Test-Research Edition*. New York: The Psychological Corporation.

Farmer, K., Cady, R., Bleiberg, J., & Reeves, D. (2000). A pilot study to measure cognitive efficiency during migraine. *Headache*, *40*(8), 657–661.

Farmer, K., Cady, R., Bleiberg, J., Reeves, D., Putman, G., O'Quinn, S., et al. (2001). Sumatriptan nasal spray and cognitive function during migraine: Results of an open-label study. *Headache*, *41*, 377–384.

Fischer, J. S. (2001). Cognitive impairment in multiple sclerosis. In S. D. Cook (Ed.), *Handbook of multiple sclerosis* (3rd ed., pp. 233–255). New York: Marcel Dekker.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental State: A practical method for grading the mental state of patients for the clinician. *Journal of Psychiatric Residency*, *12*, 189–198.

Fulton, J. C., Grossman, R. I., Udupa, J., Mannon, L. J., Grossman, M., Wei, L., et al. (1999). MR lesion load and cognitive function in patients with relapsing-remitting multiple sclerosis. *AJNR. American Journal of Neuroradiology*, *20*(10), 1951–1955.

Futrell, N., Schultz, L. R., & Millikan, C. (1992). Central nervous system disease in patients with systemic lupus erythematosus. *Neurology*, *42*(9), 1649–1657.

Gibb, W. R. (1992). Neuropathology of Parkinson's disease and related syndromes. In J. M. Cedarbaum, & S. T. Gancher (Eds.), *Parkinson's disease* (pp. 361–376). Philadelphia: W.B. Saunders.

Ginsburg, K. S., Wright, E. A., Larson, M. G., Fossel, A. H., Albert, M., Schur, P. H., et al. (1992). A controlled study of the prevalence of cognitive dysfunction in randomly selected patients with systemic lupus erythematosus. *Arthritis and Rheumatism*, *35*(7), 776–782.

Gonzalez, C. F., Swirsky-Sacchetti, T., Mitchell, D., Lublin, F. D., Knobler, R. L., & Ehrlich, S. M. (1994). Distributional patterns of multiple sclerosis brain lesions, magnetic resonance imaging—Clinical correlation. *Journal of Neuroimaging*, *4*(4), 188–195.

Goodglass, H., & Kaplan, E. (1996). *The assessment of aphasia and related disorders* (2nd ed.). Media, PA: Williams and Wilken.

Gronwall, D. (1977). Paced Auditory Serial Adition task: A measure of recovery from concussion. *Perceptual and Motor Skills*, *44*, 367–373.

Hanly, J. G., Fisk, J. D., Sherwood, G., Jones, E., Jones, J. V., & Eastwood, B. (1992). Cognitive impairment in patients with systemic lupus erythematosus. *Journal of Rheumatology*, *19*(4), 562–567.

Hanly, J. G., Hong, C., Smith, S., & Fisk, J. D. (1999). A prospective analysis of cognitive function and anticardiolipin antibodies in systemic lupus erythematosus. *Arthritis and Rheumatism*, *42*(4), 728–734.

Hay, E. M., Black, D., Huddy, A., Creed, F., Tomenson, B., Bernstein, R. M., et al. (1992). Psychiatric disorder and cognitive impairment in systemic lupus erythematosus. *Arthritis and Rheumatism*, *35*(4), 411–416.

Hay, E. M., Huddy, A., Black, D., Mbaya, P., Tomenson, B., Bernstein, R. M., et al. (1994). A prospective study of psychiatric disorder and cognitive function in systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, *53*(5), 298–303.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *WCST manual revised & expanded*. Odessa, FL: PAR, Inc.

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, Inc.

Holliday, S., Navarrete, M., Hermosillo-Romo, D., Valdez, C., Saklad, A., Escalante, A., et al. (2003). Validating a computerized neuropsychological test battery for mixed ethnic lupus patients. *Lupus*, *12*(9), 697–703.

Ingraham, L., & Aiken, C. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology*, *10*, 120–124.

Johnson, S. K., DeLuca, J., Diamond, B. J., & Natelson, B. H. (1996). Selective impairment of auditory processing in chronic fatigue syndrome: A comparison with multiple sclerosis and healthy controls. *Perceptual and Motor Skills*, *83*(1), 51–62.

Kaell, A. T., Shetty, M., Lee, B. C., & Lockshin, M. D. (1986). The diversity of neurologic events in systemic lupus erythematosus, prospective clinical and computed tomographic classification of 82 events in 71 patients. *Archives of Neurology*, *43*(3), 273–276.

Kane, R., Short, P., Sipes, W., & Flynn, C. (2005). Development and validation of the Spaceflight Cognitive Assessment Tool for Windows (WinSCAT). *Aviation, Space, and Environmental Medicine*, *76*(6, Section II), B183–B191.

Koller, W. C., & Montgomery, E. B. (1997). Issues in the early diagnosis of Parkinson's disease. *Neurology*, *49*(Suppl. 1), S10–S25.

Leritz, E., Brandt, J., Minor, M., Reis-Jensen, F., & Petri, M. (2000). Subcortical cognitive impairment in patients with systemic lupus erythematosus. *Journal of the International Neuropsychological Society: JINS*, *6*(7), 821–825.

Levinson, D., Reeves, D., Watson, J., & Harrison, M. (2005). Automated neuropsychological assessment metrics (ANAM) measures of cognitive effects of Alzheimer's disease. *Archives of Clinical Neuropsychology*, *20*, 403–408.

Levinson, D. M., & Reeves, D. L. (1997). Monitoring recovery from traumatic brain injury using automated neuropsychological assessment metrics (ANAM V1 0). *Archives of Clinical Neuropsychology*, *12*(2), 155–166.

McNicholl, J. M., Glynn, D., Mongey, A. B., Hutchinson, M., & Bresnihan, B. (1994). A prospective study of neurophysiologic, neurologic and immunologic abnormalities in systemic lupus erythematosus. *The Journal of Rheumatology*, *21*(6), 1061–1066.

Rao, S., Leo, G., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis: I. Frequency, patterns, and prediction. *Neurology*, *41*, 685–691.

Reeves, D. L., Winter, K., LaCour, S., Raynsford, K., Kay, G., Elsmore, T., et al. (1992). *Automated Neuropsychological Assessment Metrics documentation: Vol. I. Test administration guide (Technical Report)*. Office of Military Performance Assessment Technology, Silver Spring, Maryland.

Reitan, R. (1979). *Manual for administration of neuropsychological test batteries for adults and children*. Tucson, AZ: Reitan Neuropsychological Laboratory.

Reitan, R. M. (1992). *Trail Making Test manual for administration and scoring*. Tucson, AZ: Reitan Neuropsychological Laboratory.

Roebuck-Spencer, T., Bleiberg, J., Nowak, M., Lapteva, L., Yarboro, C., Weickert, T., et al. (2004, July). Use of computerized testing to predict neuropsychological status in lupus. In *Paper presented at the Annual Meeting of the American Psychological Association*.

Roebuck-Spencer, T. M., Yarboro, C., Nowak, M., Takada, K., Jacobs, G., Lapteva, L., et al. (2006). Use of computerized assessment to predict neuropsychological functioning and emotional distress in patients with systemic lupus erythematosus. *Arthritis & Rheum*, *55*(3), 434–441.

Soukup, V. M., & Adams, R. L. (1996). Parkinson's disease. In R. L. Adams, O. A. Parsons, J. L. Culbertson, & S. J. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders* (pp. 243–267). Washington, DC: American Psychological Association.

Stevens, J. (1996). *Applied multivariate statistics*. Mahway, NJ: Earlbaum.

van den Burg, W., van Zomeren, A. H., Minderhoud, J. M., Prange, A. J., & Meijer, N. S. (1987). Cognitive impairment in patients with multiple sclerosis and mild physical disability. *Archives of Neurology*, *44*(5), 494–501.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-Third Edition: Administration and scoring manual*. San Antonio: The Psychological Corporation.

Wilken, J. A., Kane, R., Sullivan, C. L., Wallin, M., Usiskin, J. B., Quig, M. E., et al. (2003). The utility of computerized neuropsychological assessment of cognitive dysfunction in patients with relapsing-remitting multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)*, *9*(2), 119–127.

Wishart, H., & Sharpe, D. (1997). Neuropsychological aspects of multiple sclerosis: A quantitative review. *Journal of Clinical and Experimental Neuropsychology*, *19*, 810–824.