The Structure and Meaning of a Computerized Neurocognitive Test Battery

C. Thomas Gualtieri[*1], Aaron S. Hervey[2]

[1,2]Departments of Neuropsychiatry & Neuropsychology, North Carolina Neuropsychiatry, PA,
1829 E. Franklin St.,
Chapel Hill, NC
27514,
United States
[*1]TG@ncneuropsych.com; [2]AHervey@ncneuropsych.com

***Abstract*-Background: There exists a prevailing assumption that a neurocognitive test administered by a computer is equivalent to the same test administered conventionally by a human being. There is reason to question this assumption.**

Method & subjects: A computerized neurocognitive test (the CNT) was subjected to critical analysis. Study 1 addressed the appropriate standardization method and was conducted in 3420 normal subjects between the ages of 4 and 94. Study 2 examined the factor structure of the CNT in the same group. Study 3 examined the discriminant powers of the CNT in comparison with 3295 normal subjects compared to 4084 subjects with attention deficit hyperactivity disorder (ADHD), 694 patients with severe traumatic brain injuries and 90 patients with mild cognitive impairment (MCI) or early dementia. Study 4 was a comparison of the CNT with The Wechsler Adult Intelligence Scale (WAIS) in a clinical sample of 179 patients.

Analysis & results: Study 1: Regression analysis indicated that age and years of education contributed significantly to a subject's performance on the CNT but race and gender did not. Study 2: Exploratory and confirmatory factor analyses indicated that the seven tests of the CNT generated three factors: memory, processing speed and attention. Processing speed was the latent variable underlying performance on the test battery. Study 3: Neither stepwise discriminant function analysis nor logistic regression indicated a specific pattern to performance for the three clinical groups relative to each other. Study 4: Performance on the CNT was most highly correlated with the most highly g-loaded WAIS subtests. Two tests of the CNT - visual memory and shifting attention - were predictive of Full Scale Intelligence Quotient (FSIQ).

Conclusion: These four studies raise questions about the continued use of computerized tests in clinical practice and research. The latent variable underlying test performance is processing speed. On the CNT, tests of executive function, memory and attention are, in the main, measuring the processing speed component of those functions. A critical examination of one CNT suggests that it may be doing less than it was designed to do.

Keywords- Computerized Test; Test Standardization; Processing Speed; General Mental Ability

I. INTRODUCTION

In theory, cognitive tests should be amenable to computerization; in fact, many have already been computerized. Computerized tests are widely used, although in circumscribed areas: "performance assessment batteries" like the National Evaluation Series (NES) have been used for years in toxicology research [1]; the Automated Neuropsychological Assessment Metrics (ANAM) is used by the American military for the early diagnosis of brain injury [2]; ImPact is a concussion screening instrument administered by athletic trainers in high schools and colleges [3]. The pharmaceutical industry employs a number of different computerized neurocognitive tests in clinical trials. The advantages of computerized testing in such endeavors have been enumerated and hardly bear repeating.

The assumption behind all of this is that a test administered by a computer is equivalent to the same test traditionally administered by a human being. As long as the two are similar in construction, the computerized test is said to have construct or content validity. If it generates different scores when normal subjects are compared to people with a pathological condition, then it has discriminant validity. If the test meets these two requirements, then it is appropriate to use in the clinic, in the locker-room, in a battlefield hospital or wherever.

When these assumptions are examined more closely, however, one finds that a given test administered by a computer is not quite the same as the same test administered in the conventional way. Recent studies, for example, of neuropsychological tests that seem quite suitable for computerization indicate that computer-based versions of the line orientation test, the enhanced cued recall test and the Stroop test generated significant differences from the results of the same tests when they were administered conventionally [4, 5]. Such results are counter-intuitive; correlations between mental tests should be high if the tests are equivalent even if the modality of stimulus presentation is different. An example of this is verbal presentation by an examiner and visual presentation by the computer. Or when the response modality is different, for example, word recall in response to an examiner versus word recognition in response to a computer [1]. Correlations, however, do not indicate equivalence.

Although there are a host of papers concerned with applications for computerized tests, including several we have written ourselves, there is not much in the way of critical analysis of the results they generate. Therefore, in 2012, representatives of the American Academy of Clinical Neuropsychology expressed concerns over the appropriate psychometric development of “computerized neuropsychological assessment devices” (CNADs). They raised other pertinent issues, such as demographic factors that may affect examinee performance and the need for checks on response validity and effort for the CNAD [6].

We describe four studies concerned with the psychometric properties of a CNAD, referred to as the “computerized neurocognitive test” (CNT). These four studies examine: (1) expected performance in normal subjects and the relative effects of age, education, gender and race; (2) the structure of the test defined by factor analysis; (3) the discriminant validity of the test in four patient groups; and (4) the relation of CNT performance to a general mental ability test, the Wechsler Adult Intelligence Scale (WAIS).

II. METHODS

Four studies are described. The studies and the analytic methods are described in the relevant sections. The subjects were normal volunteers who participated in the standardization studies of the CNT or patients who had taken the CNT as part of a comprehensive neuropsychiatric examination. The clinical subjects were a “convenience sample” of patients tested with the CNT in neuropsychiatry clinics (Table 1).

TABLE 1 SUBJECTS IN THE FOUR STUDIES

Study	N		Age Range	Mean Age	% Male	Mean Education	% Non-White
1.Standardization	3420	Normal Volunteers	4-90	39.6	51%	12.6	8%
2.Factor Analysis	3420	Normal Volunteers	4-90	39.6	51%	12.6	8%
3.Discriminant Validity							
Normal Subjects	3295	Normal Volunteers	4-94	39.09	39.9%	13.69	8.3%
ADHD (ADD)	4084	Patients	4-77	18.87	62.1%	8.92	16.4%
TBI	694	Patients	7-85	39.59	69.6%	12.79	7.6%
Early Dementia	90	Patients	19-88	69.1	44.4%	15.3	4.4%
4.WAIS Comparison	179	Patients	13-80	27.6	60.9	13.09	17.9

The CNT battery is an updated version of a computerized test battery called CNS Vital Signs (Central Nervous System Vital Signs), developed by the author (CTG) and introduced in 2003 [7]. CNS Vital Signs is currently used by clinicians and researchers and has been applied in studies of patients with ADHD [8], traumatic brain injury [9], mild cognitive impairment [10], mood disorders [11] and other clinical conditions [12]. The CNT is identical to the original test battery, save these differences: standardization and scoring have been changed in accord with the results of this paper; validity measures are incorporated as described in a companion paper; the new test is internet-based; and it is not a commercial product.

The CNT battery contains eight tests that generate nine scores. Seven of these tests are the topic of this paper. However the eighth, keyboard speed, is a new test that is still in the development phase. It was introduced as an additional validity measure. The seven tests were originally chosen because they were thought to address distinct cognitive domains (Table 2).

TABLE 2 THE COMPUTERIZED TEST BATTERY (CNT)

Test	Abbreviation	Test Time (mins)	Factor
Verbal Memory	VBM	3	Memory
Visual Memory	VIM	3	
Finger Tapping	FTT	3	Motor Speed and Coordination
Symbol Digit Coding	SDC	4	Central Processing Speed
Shifting Attention Test	SAT	3	
Stroop Test	ST	5	
	RT		Sustained Attention
Continuous Performance Test	CPT	6	

The verbal memory (VBM) and visual memory (VIM) tests are adaptations of the Rey Auditory Verbal Learning Test and the Rey Visual Design Learning Test [13, 14]. VBM and VIM are tests of recognition memory; they are administered at the beginning and the end of the battery, yielding scores for immediate and delayed memory. The finger tapping test (FTT) is administered in three 10 second segments to each hand. The symbol digit coding (SDC) test is based on the symbol digit modalities test [15]. The Stroop Test (ST) has three parts that generate simple and complex reaction times [16]. A “response time” (RT) score is generated by averaging the two complex reaction time scores from the Stroop test. The ST also generates an error score. The Shifting Attention Test (SAT) measures the subject’s ability to shift from one instruction set to another quickly and accurately. Other computerized batteries, like the NES2 (Neurobehavorial Evaluation System 2), Cogstate and CANTAB (Cambridge Neuropsychological Test Automated Battery) include shifting attention tests. Color-shape tests like the SAT have also been used in cognitive imaging studies [17, 18]. The SAT score is calculated by subtracting the number of errors from the number of correct responses. The Continuous Performance Test presents 40 targets (the letter “B”) embedded among 160 non-target letters over a five-minute interval [19].

The tests generate raw scores and standard scores. Raw scores were used in these studies, except where otherwise indicated. Scores are standardized by adjusting for age and education level, as per data presented below.

The CNT normative database includes 4400 normal individuals. Normal subjects were in good health without past or present psychiatric or neurological diagnoses and free of centrally active medications. They signed written informed consent for the use of their de-identified data for test standardization. From that number, the data of 3,420 subjects was used in this normative analysis. Their mean level of education was significantly higher than national averages because of the demographics of the community to which they belonged; therefore, subjects were randomly selected such that participants with higher levels of education were less likely to be included in the analyses. The final group had a mean education level of 12.6 years, which matches the mean education level of the US population as per the 2010 census. The subjects ranged in age from 4 to 90 years; 4, 5 and 6 year olds who were able to take the test were precocious children and the normed values do not apply to those ages.

III. TESTS

A. Standardization

Psychological tests are ordinarily standardized by age. Other elements that may contribute to variance in test scores are levels of education, gender and race. The following analysis was done to determine the relative importance of those three elements. We find that age contributes the most to the variance in CNT scores; education is also salient, but gender and race are not relevant.

1) Analysis, Results

Years of education were taken from subjects' self-reports. Race also was self-reported and based on US census categories. Tables 3 and 4 show regression coefficients for the linear prediction of education, gender, and race, respectively. Both raw and standardized coefficients are given, as well as their probability values. Probabilities of less than .05 are considered significant.

TABLE 3 THE IMPACT OF EDUCATION LEVELS ON 8 TESTS IN THE CNT BATTERY

Test	Education			Gender			Race		
	Estimate	Standardized	P	Estimate	Standardized	P	Estimate	Standardized	P
VBM	.113	.074	.175	-.274	-.020	.49	-1.496	-.065	.025
VIM	.086	.062	.21	.7965	.069	.01	-1.106	-.058	.043
SDC	.880	.275	<.0001	-2.349	-.077	.0015	-1.642	-.033	.1817
Stroop RT	-5.751	-.226	<0001	9.217	.038	.137	14.767	.036	.16
SAT	1.04	.253	<.0001	-.699	-.019	.46	-3.305	-.054	.0357
Stroop Errors	-.072	-.098	.1394	.258	.037	.27	.380	.021	.53
CPT	.024	.108	.03	-.189	-.043	.11	.415	-.057	.036
FTT	.607	.135	.004	5.757	.148	<.0001	-1.204	-.019	.49

Note: Table 3. VBM: verbal memory test; VIM: visual memory test; SDC: symbol digit coding test; Stroop RT: average reaction time on parts 2 and 3 of the Stroop test; CPT: continuous performance test; FTT: finger tapping test.

TABLE 4 AGE AND EDUCATION ARE RELEVANT TO THE NORMATIVE VALUES OF THE CNT

Test	Intercept	Age	Age 2	Education
VBM	48.49**	1.34*	-.185**	.10
VIM	44.84**	.819*	-.146**	.071
SDC	26.46**	9.42**	-1.14**	.838**
Stroop RT	845.48**	-65.61**	8.06**	-5.43**
SAT	7.53**	10.77**	-1.12**	.957**
CPT	99.59**	-3.25**	.213*	.71**
Stroop Errors	5.70**	-1.05**	.10**	-.07
FTT	82.71**	11.26**	-1.16**	.53*

*Note: Table 4. **p<.01; *p<.05. Note, Table 3. VBM: verbal memory test; VIM: visual memory test; SDC: symbol digit coding test; Stroop RT: average reaction time on parts 2 and 3 of the Stroop test; CPT: continuous performance test; FTT: finger tapping test.*

Education is a significant predictor of performance in all of the tests except verbal and visual memory, and Stroop test errors. Based on standardized coefficients, symbol digit coding, Stroop response time, and the shifting attention test are most affected by education. In the case of gender, only finger tapping is statistically significant and shows a beta value higher than 0.1. None of the racial differences generate meaningful betas.

In order to develop an appropriate set of norms for the CNT, a quadratic function for age, and years of education was used to predict performance on the eight tests (Table 4).

B. Factor Analysis

The eight tests of the CNT were selected because they were believed to address discrete cognitive domains. That assumption was tested by factor analysis of the data from the 3,420 normal subjects described in the preceding section.

1) Analysis, Results

Both exploratory and confirmatory factor analyses were conducted. A first set of exploratory factor analysis (EFA) was conducted initially. However, in order to avoid potential problems due to random aspects of the sample (see [20] for a description of the dangers of conducting an EFA) the EFA was done again on a random sample that consisted of roughly 50 percent of the data. Then, the results of the EFA were used as a confirmatory factor analysis (CFA) on the other random half of the data set in order to cross-validate the data. All factor analyses were conducted using SPSS 17.0 (Statistical Package for the Social Sciences) and AMOS 20.0 (Analysis of Moment Structures).

The structure of the CNT is best described by a three-factor mode. The first is a memory factor that includes the results of the verbal and visual memory tests. The second is a processing speed factor, including the symbol digit coding and shifting attention tests and response time on the Stroop test. The third, and last, is an attention factor, including the continuous performance test and Stroop test errors. The FTT does not load with any of these three factors. This model is shown below in Fig. 1 with correlation coefficients and is an adequate fit to the data (RMSEA=.065; CFI=.961).

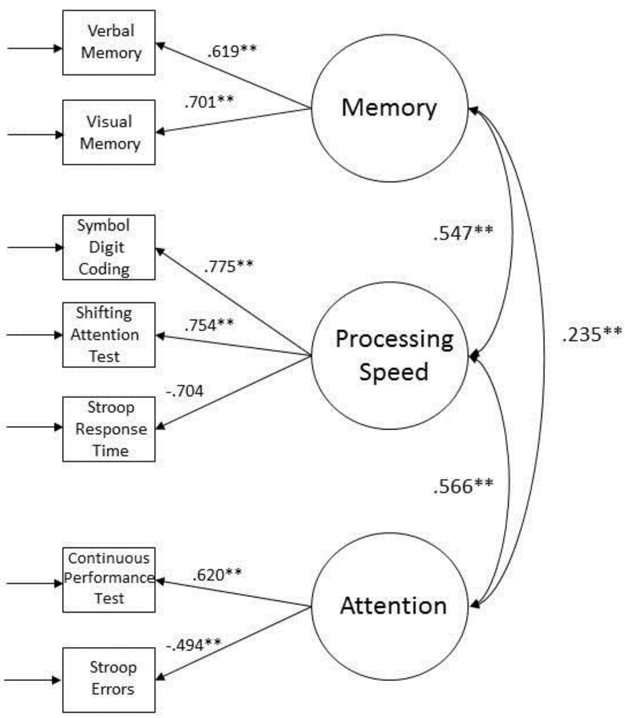

Fig. 1 Factor structure of the seven tests

C. Discriminant Validity

If patients with different cognitive disorders perform differently on a CNAD, the test is said to have discriminant validity. If a CNAD is a broad-spectrum test, it should generate differences in degree – higher scores in group A, lower scores in group B – and also different patterns of response. Three patient groups were compared to normal subjects: adult patients with ADHD, traumatic brain injury (TBI) and early dementia.

1) Subjects

The CNT database contains records of 12,400 subjects with various neurocognitive disorders. From the database, three clinical groups were chosen:

- ADHD group: adult patients (=> 18 years) diagnosed with attention-deficit disorder (ADD) residual type, by DSM-IVtr criteria, with no concurrent psychiatric or medical conditions and medication free.
- TBI group: adult patients who had sustained severe traumatic brain injuries (GCS < 8, LOC > 24 hrs) but had made good recoveries, were living independently and capable of taking the CNT unassisted. (The CNT requires a 4th grade reading level.)
- Early dementia group: patients who were clinically diagnosed by a neuropsychiatrist or neuropsychologist in our clinic and scored 1 on the clinical dementia rating scale, but who were still living independently.

2) Analysis, Results

The three clinical groups were compared to normal adults from the normative sample (Fig. 2). Both multivariate analysis of variance (MANOVA) (controlling for age, race, gender and education) and stepwise discriminant function analysis were conducted to test whether or not the CNT could discriminate between normal, ADD, traumatic brain injury (TBI) and early dementia patients (DEM). When the results For the MANOVA proved to be significant further pairwise post hoc tests were conducted. Stepwise discriminant function analysis and logistic regression were also performed comparing each group to another in pairwise fashion, to determine if a specific pattern could be discerned between the groups.

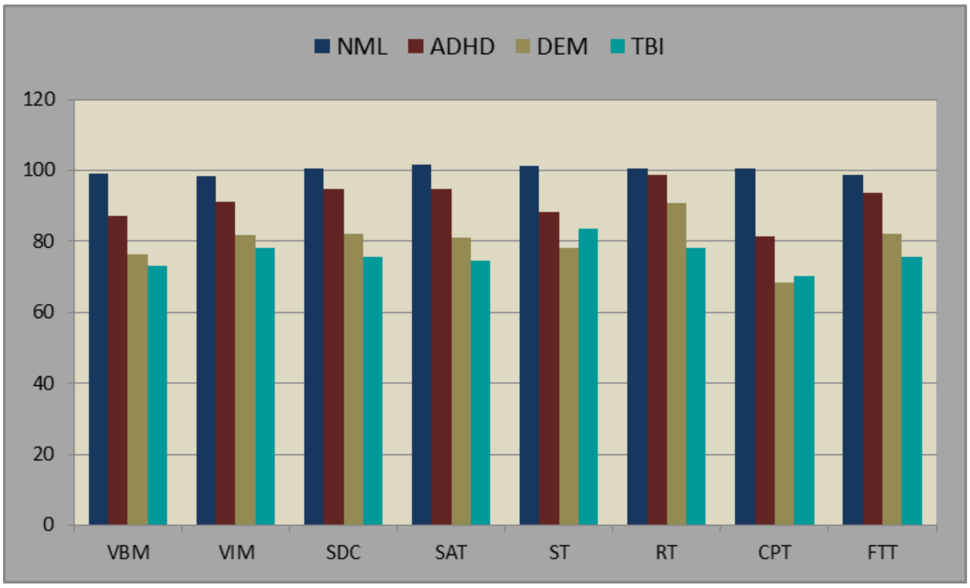

Fig. 2 Test scores for four groups

Note: Fig. 2. VBM: verbal memory test; VIM: visual memory test; SDC: symbol digit coding test; Stroop RT: average reaction time on parts 2 and 3 of the Stroop test; ST: Stroop test errors; CPT: continuous performance test; FTT: finger tapping test.

The MANOVA indicated significant differences among the four groups on all of the tests. The post hoc test (Tukey's honest significant differences) was applied to control for type 2 errors. Applying this correction, all of the group differences were significant with the exception of the following: on the VBM test, there were no significant group differences between normal subjects and ADD subjects; the ADD, TBI, and dementia subjects were not different in their performance on the continuous performance test (CPT). Neither stepwise discrimination nor logistic regression was able to identify a specific pattern of response to identify the four groups. There were overall differences among the groups based on the discriminant function analysis, and four tests (VBM, ST, SAT and FTT) were sufficient to separate the groups. On an overall level, ADD patients were expected to score an average of .125 units below normal patients; TBI patients were expected to score 1.05 units below normal; and dementia patients were expected to score 2.5 units below normal patients.

D. Relation to General Mental Ability

An appropriate question to raise is the relationship of performance on the CNT to general mental ability as measured by the Wechsler Adult Intelligence Scales (WAIS).

1) Subjects

A convenience sample of 179 patients was identified who had taken the CNT along with a psychological evaluation that included the Wechsler Adult Intelligence Scales (WAIS III or IV). The patients were evaluated for ADD, learning disability or mild cognitive impairment.

2) Analysis, Results

This analysis, in contrast to the others, above, utilized standard scores for the eight tests and included four additional measures. The Pearson Product-moment correlation results are shown in Table 5. The correlations between the CNT scores and the WAIS were almost all significant but most were very low. The components of the WAIS that correlated most highly with the CNT were FSIQ > VIQ > PIQ, and among the subtests, vocabulary, similarities, block design and matrix reasoning. The components of the CNT that were most highly correlated with the WAIS were VBM, VIM, SDC and SAT.

Multiple regression (controlling for age, gender, and education) was run to determine which or if any tests could significantly predict full-scale IQ (FSIQ) (Table 6). The overall model generated by multiple regression indicated that the only significant predictors of FSIQ are VIM and SAT (F(7,126)=2.959, p=.007, coefficient of determination R=.099). The results are given in Table 5. The most g-loaded tests are VIM and SAT, which correlated with the FSIQ at r = 0.52 and at r =0.59

respectively. When the VIM and SAT standard scores were averaged to produce a composite VIM-SAT score, the correlation for the FSIQ was 0.64. It is interesting to note that the coding test on the CNT correlated with digit symbol coding at r = 0.10.

TABLE 5 CORRELATIONS OF THE CNT AND THE WAIS

	VBMss	VIMss	SDCss	RTss	SATss	STss	CPTss	FTTss
FSIQ	.410**	.516**	.465**	.168*	.590**	0.087	.231**	.229**
VIQ	.534**	.525**	.417**	0.044	.408*	0.097	0.136	0.069
PIQ	0.283	.443**	.340*	-0.059	.375*	0.096	.289**	0.158
VC	0.19	0.305	0.082	0.021	0.141	0.018	0.13	0.037
PO	-0.03	0.18	-0.327	0.01	0.08	0.185	.272**	0.103
WM	0.309	-0.108	-0.097	0.012	-0.164	0.194	-0.041	0.013
PS	0.214	-0.007	-0.179	0.093	-0.088	0.029	0.059	0.03
PC	0.097	-0.003	0.093	0.058	0.287	0.05	.260*	.249**
VOC	.354**	.512**	.425**	0.088	.570**	0.038	.177*	.167*
DSC	0.097	0.073	0.103	-0.01	0.036	-0.059	0.089	0.055
SIM	.513**	.571**	.408**	0.1	.396*	-0.002	0.186	.228*
BD	0.303	.365*	.392*	0.115	.434**	0.067	.262**	0.053
ARI	0.336	0.157	0.21	-0.093	0.153	0.097	0.094	-0.064
MR	.269*	.424**	.458**	0.098	.509**	0.134	.222**	.163*
DS	0.352	-0.023	0.003	0.106	0.172	0.138	0.031	0.098
INF	0.267	0.158	0.09	-0.075	0.146	0.051	0.031	-0.111
PA	-0.03	0.046	-0.137	-.232*	-0.187	0.074	0.004	0.072
COM	0.163	0.357	.429*	0.01	0.309	.247*	0.146	-0.049
SS	0.122	-0.062	0.105	0.136	-0.233	0.073	0.049	0.082
LNS	0.013	-0.22	-0.161	0.062	-0.268	0.144	-0.082	0.098

Note: Table 5. VBM: verbal memory test; VIM: visual memory test; SDC: symbol digit coding test; Stroop RT: average reaction time on parts 2 and 3 of the Stroop test; CPT:continuous performance test; FTT: finger tapping test; "ss" indicates standard scores; FSIQ: full scale intelligence quotient; VIQ: verbal intelligence quotient (WAIS 3); PIQ: performance intelligence quotient; VC: verbal comprehension index (WAIS 4); PO: perceptual organization index; PS: processing speed index; WM: working memory index; PC: picture completion; VOC: vocabulary; DSC: digit symbol coding; SIM: similarities; BD: block design; ARI: arithmetic; MR: matrix reasoning; DS: digit search; INF: information; PA: picture arrangement; COM: comprehension; SS: symbol search; LNS: letter number sequencing.

TABLE 6 REGRESSION ON FSIQ

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Standard Error	Beta		
1	(Constant)	58.869	21.271		2.768	.007
	VBM	-.036	.219	-.016	-.162	.871
	VIM	.453	.226	.187	2.003	.047
	RT	.001	.008	.008	.078	.938
	SDC	-.008	.076	-.010	-.104	.918
	SAT	.170	.065	.245	2.618	.010
	CPT	.768	.433	.164	1.772	.079
	FTT	-.016	.053	-.029	-.304	.761

Note: Table 6. VBM: verbal memory test; VIM: visual memory test; SDC: symbol digit coding test; Stroop RT: average reaction time on parts 2 and 3 of the Stroop test; CPT: continuous performance test; FTT: finger tapping test.

IV. CONCLUSIONS AND SUMMARY

In an earlier publication we described the test-retest reliability of the CNT and the correlations between the CNT and equivalent neuropsychological tests administered conventionally [7]. In the companion paper, we describe a method for establishing the validity of a CNT test session by comparing the results of normal subjects to patients with brain injuries, directed malingerers and a group of patients who were thought to be manipulating their test results. In this paper we address two additional practical issues in two studies: how the tests should be standardized, in Study 1; and how the tests should be scored, in Study 2.

With respect to standardization, most if not all computerized tests employ normative data relative to subjects' age. In Study 1 we establish that not only age but education levels are also relevant to the analysis of test performance. Race and gender, however, are not. The practical implications of this outcome are illustrated in Fig. 3, which shows educated-adjusted mean values on the coding test for individuals with ten years of education compared to those with 20.

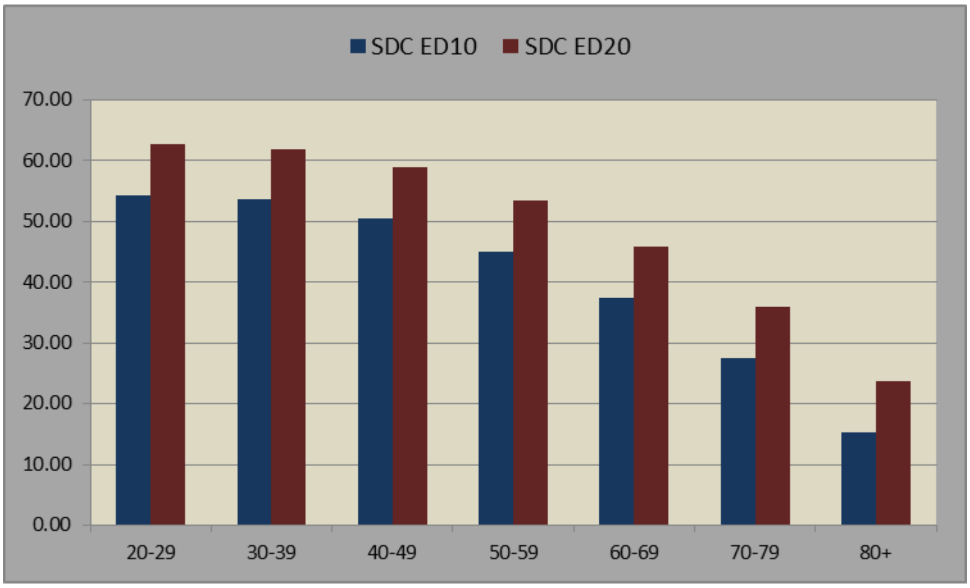

Fig. 3 SDC norms relative to years of education

Note: Fig. 3. SDC: symbol digit coding test; ED10: subjects with ten years of education; ED 20: subjects with 20 years of education; y axis: SDC raw score; x axis: age group.

Study 2 addresses a more fundamental question about the CNT, a practical question with theoretical implications: what, precisely, does it measure? The rationale for the tests in the CNT was to measure distinct domains of cognitive ability: verbal and visual memory, motor speed and coordination (FTT), processing speed (SDC), sustained attention (CPT) and executive functions like inhibition/disinhibition (the Stroop test) and cognitive flexibility (SAT). Factor analysis, however, indicates that only three broad areas of cognitive function are measured: memory (VBM and VIM), processing speed (SDC, ST, SDC) and effortful attention (Stroop errors and the CPT).

The CNT has a measure of specificity. Three distinct factors are elicited. The CNT is able to discriminate two aspects of performance on the ST: one is the ST response time, which loads on the processing speed factor, and ST errors load on the attention factor. The FTT does not load on any of the three factors. That is reassuring, because the CNT generates scores when a subject presses a button on a keyboard. The fact that the FTT does not load with the processing speed factor indicates that the speed dimension that emerges as a central factor in the CNT is not simple motor speed; it is central information processing speed.

Although the CNT is comprised of three cognitive factors, it is likely that the latent variable underlying performance on the CNT is processing speed. The processing speed factor is strongly correlated with the memory and attention factors, while the memory and attention factors are only weakly correlated with each other. The SDC, ST and SAT are, ostensibly, three different tests but no less than 50 percent of the variance in performance on these tests is explained by a single factor. This factor, which we refer to as processing speed, also accounts for 30 percent of variance in the memory factor and 32 percent in the attention factor.

If processing speed is a superordinate domain that governs performance on all the tests of the CNT, it may be a function of the time constraints imposed by the computer. In order to make the CNT a quick and efficient test, subjects have only a limited time to respond to each item on every test. Thus, subjects who can respond quickly perform well, and subjects who respond slowly are penalized. The CNT does measure memory and attention, but the aspects of memory and attention that it captures are in the context of the celerity with which a subject processes and executes a task. Processing speed, therefore, seems to be the function that underlies performance on all the tests of the CNT, to a greater or lesser extent; the more articulated dimensions addressed by such tests are more accessible when they are administered with fewer time constraints and by a skilled examiner.

If the CNT were a broad-spectrum or a comprehensive test battery, one would expect to find different *patterns* of response when different patient groups are compared. If, on the other hand, central processing speed is at the heart of the test battery, its flexibility with respect to distinguishing different conditions will be limited. In that light, the results of Study 3 are pertinent. It is evident that the CNT can discriminate among normal subjects and patients with three cognitive disorders. However, the nature of the discrimination is a severity gradient based on all the tests, from normal subjects to ADD patients to mild dementia patients to TBI patients. Although ADHD, TBI and early dementia are very different conditions, they do not generate different patterns of response on the CNT. Rather, differences among the four groups occur to the same degree in virtually all of the tests. Pairwise comparisons of eight test scores among four groups revealed only one non-significant difference: TBI and early dementia patients perform equally poorly on the CPT.

If the latent variable underlying performance on the CNT is central processing speed, one should expect to see a meaningful relationship with a test of general mental ability; in study 4, the latter was represented by the WAIS-III. There is ample support for this hypothesis: Spearman's theory that complex problem-solving tasks necessarily share some specific information-processing resource; the contention of Salthouse that speed of processing is a fundamental part of the "architecture of the cognitive system" as it develops across the lifespan [21, 22]. Jensen, too, is associated with a theory that complex reaction time measures are an index of the speed and efficiency with which the nervous system processes elementary information. Inspection time, for example, is said to be "an information-processing correlate of psychometric intelligence" [23] and meta-analysis of inspection time (IT) studies indicate correlations with IQ at r= -0.51 - 0.54 [24, 25], an association that is strengthened by analogous studies in behavioral genetics [26]. Studies employing diverse tasks, like choice reaction time, IT, Sternberg and Posner paradigms, reading rates, and coding of numbers or letters, show an average relationship between psychometric intelligence and speed of information processing of about r=0.30 for single elementary cognitive tasks, which rise to multiple relationships up to .50 or .60 when combining several such tasks [25, 27, 28, 29].

When we compared the CNT to the WAIS in adult patients, the highest associations between the WAIS and the CNT occurred in relation to the g-loading of the WAIS subtests: vocabulary, matrix reasoning, similarities and block design. FSIQ is moderately correlated with the two purest measures of processing speed on the CNT, SDC (r= 0.47) and SAT (0.59). The association between the two memory tests and FSIQ is also explicable at least in part in terms of complex problem-solving. This means the ability to generate an effective strategy to remember words and abstract figures within a very short span of time (2 seconds). To develop such a strategy to remember words is less difficult than to remember abstract figures, and this may account for the higher correlation between VIM and FSIQ. A derived score, VIM-SAT, the average of the VIM and SAT standard scores, correlates with FSIQ at r = 0.644.

There are some anomalies in the CNT/WAIS comparison, however. For example, one notes that neither the processing speed (PS) factor of the WAIS nor the coding test are highly correlated with any of the scores on the CNT and, in particular, the CNT PS factor and the SDC. Further, WAIS measures that contribute to the VIQ score are more closely related to CNT scores than the subtests that contribute to PIQ, and the vocabulary subtest is more closely related than matrix reasoning. If the CNT were only a measure of central processing speed, one should expect the opposite to be the case. Although only so much can be made of the results of a correlation analysis, the latter anomaly affirms the importance of education to CNT performance.

The conclusions we draw from these four studies are that the CNT is, indeed, a cognitive test that is related in a meaningful way to age and education level, to general mental ability, to central processing speed and to the severity of a patient's cognitive condition. It is not a measure of seven neuropsychological domains, as it was designed to be. Nor is the CNT a diagnostic test. If one knows what a patient's diagnosis is, the CNT indicates the degree of cognitive weakness that characterizes that individual. But it is not possible to examine CNT results and say, "*These results indicate ADHD*" or indeed any other clinical condition. The results may be consistent with ADHD, they may indicate that a patient's cognitive weaknesses are in excess of what one ordinarily sees in cases of ADHD, but (as one reads in many test reports) *clinical correlation is needed*. This, to our knowledge, is a new finding. Existing CNAD literature invariably presents the results of comparisons between normal subjects and patients in a known group, for example, concussion patients or workers who were exposed to lead or other toxins. These are taken to be indicators of discriminant validity, and they probably are but only within a limited frame of reference. We are not aware of any reports in the current store of CNAD literature that indicate reproducible and, more important, specific differences between patient groups.

Two questions remain to be addressed. First, are the results of these studies germane to other CNADs, or is the CNT a special case? Second: What is the CNT good for? That is, what are its proper clinical applications? The studies described here, however, may not be conducive to broad generalizations. For example, the normative database of the CNT is sound, but Hispanics are under-represented. There were substantial demographic disparities among the patient groups. The respondents represented a "convenience sample" of patients at neuropsychiatry clinics, subject to the well-known problem of referral bias, and the diagnostic procedures were clinical and not up to the standards of a clinical research unit. The WAIS subjects were mostly young people who were being evaluated for ADHD or learning disability. Only three diagnostic groups were studied; a test like the CNT could conceivably be useful in distinguishing other clinical conditions.

In spite of these shortcomings, our results are probably relevant to CNADs in general because almost all of them use the same or very similar tests and are subject to the same time constraints. But even in a widely used CNAD, the CANTAB, that is different from other CNADs in its choice of tests, seems to have the same association with general mental ability and central processing speed as the CNT. A study of the CANTAB in older subjects indicated two factors that accounted for over 60% of the variance. Factor 1 was equated with general learning and memory ability and loaded significantly with intelligence scores; factor 2 was related to speed of responding and loaded most heavily with age. Even though the CANTAB is an idiosyncratic test compared to most other CNADs, it appears to tap the same general cognitive resources as the CNT [30].

In a previous publication, we described the correlation between the tests of the CNT and equivalent conventional tests. The relevant correlations were deemed to be "moderate" but, in fact, ranged from r=0.13 to 0.79 [7]. Comparable studies of other CNADs report such correlations as "low to moderate" [1], "modest" [31, 32], "high" [33, 34] or "difficult to assess" [35]. The literature does not support the contention that any CNAD is better than another with respect to concurrent validity. Our feeling is that it is, indeed, difficult to assess the precise relation between computerized and conventional testing, but that the differences between the two approaches are at least as impressive as their similarities. The appropriate question, of course, is not whether test A correlates with test B; one of the best replicated findings in psychometrics is that, in the undamaged brain, performance on virtually any cognitive task correlates positively with performance on any other. The appropriate question is whether the two tests are measuring the same molecular components of cognitive function, rather than some superordinate principle, like *g* factor or processing speed.

The second question is: what is the CNT good for? Or, what are CNADs in general good for? The practical answer to these questions is necessarily based on our evolving theory of CNADs: that they are measures of general mental ability and central processing speed. If such is the case, and speaking just for the CNT, we have found that it is a useful gauge of the patient's functional status with respect to cognitive challenges. It may be especially pertinent to the serial evaluation of clinical conditions to which processing speed deficits are central; for example, ADHD and certain learning disabilities and clinical conditions like traumatic brain injury, multiple sclerosis and cerebrovascular white matter disease. Within a given diagnostic category it can detect subgroups of patients who have cognitive weaknesses [36]. The evaluation of a patient for diagnostic purposes, however, should never rely on the CNT, or any other CNAD. Although it was designed with that intention, the CNT is not a cyber-version of a neuropsychological test battery. However, having performed such an examination, it may be useful to have serial CNT data to follow the course of a patient's condition or to evaluate the results of treatment [37]. The CNT may also complement a conventional neuropsychological battery in identifying patients who are exaggerating their cognitive deficits, not only in medico-legal evaluations, but in the instance of patients seeking psychostimulants and who pretend to have ADHD.

Our studies of the CNT suggest that many of the clinical uses to which CNADs are being applied are appropriate. They are well suited, for example, in toxicology research, where well-defined groups of exposed and non-exposed subjects are compared; or in clinical trials, where pre- and post- measures and drug or placebo conditions can be compared. It is not appropriate, though, to assume that because study results indicate changes in memory, attention, executive function or working memory, for example, that, in fact, it is those domains that are affected. Until we have a clearer idea of what CNADs do actually measure, it is more parsimonious to assume that central processing speed and general mental ability are the source of differences that appear to be specific. The use of CNADs in concussion management in sport or on the battlefield is theoretically appropriate, although the authors, and others, have reservations about the reliance on CNADs for a number of good reasons [3, 34, 38, 39, 40]. The use of CNADs for the purpose of "dementia screening" is problematic, because they may not be very good tests of memory, executive function or visual-perceptual ability, thus missing cases of cortical disease; their sensitivity to processing speed deficits may overstate the effects of subcortical disease. CNADs are an inappropriate basis for the diagnosis of ADHD; patients with anxiety and depression or certain learning disabilities may have similar profiles.

CNADs, like the CNT, have potential as measurement tools in neurology and psychiatry and in many other medical fields. However, the study of computerized testing should not be confined to researchers with a positive bias towards the technology. A dispassionate examination may reveal that they do less than they are purported to do, at least in some respects. On the flip side, it is also entirely possible that their clinical potential is as yet unrealized. That potential is not likely to be advanced until more critical investigations are published.

REFERENCES

[1] Krengel, M., White, R. F., Diamond, R., Letz, R., Cyrus, P., and Durso, R., "A comparison of NES2 and traditional neuropsychological tests in a neurologic patient sample," *Neurotoxicology and teratology*, vol. 18(4), pp. 435-439, 1996.

[2] Coldren, R. L., Russell, M. L., Parish, R. V., Dretsch, M., and Kelly, M. P., "The ANAM lacks utility as a diagnostic or screening tool for concussion more than 10 days following injury," *Military medicine*, vol. 177(2), pp. 179-183, 2012.

[3] Mayers, L. B. and Redick, T. S., "Clinical utility of ImPACT assessment for postconcussion return-to-play counseling: psychometric issues," *Journal of clinical and experimental neuropsychology*, vol. 34(3), pp. 235-242, 2012. DOI: 10.1080/13803395.2011.630655.

[4] Aşkar, P., Altun, A., Cangöz, B., Cevik, V., Kaya, G., and Türksoy, H., "A comparison of paper-and-pencil and computerized forms of Line Orientation and Enhanced Cued Recall Tests," *Psychological reports*, vol. 110(2), pp. 383-396, 2012.

[5] Penner, I.-K., Kobel, M., Stöcklin, M., Weber, P., Opwis, K., and Calabrese, P., "The Stroop Task: Comparison Between the Original Paradigm and Computerized Versions in Children and Adults," *The Clinical neuropsychologist*, vol. 26(7), pp. 1142-1153, 2012. DOI: 10.1080/13854046.2012.713513.

[6] Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., and Naugle, R. I., "Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology," *The Clinical Neuropsychologist*, vol. 26(2), pp. 177-196, 2012. DOI: 10.1080/13854046.2012.663001.

[7] Gualtieri, C. and Johnson, L., "Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs," *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, vol. 21(7), pp. 623-643, 2006. DOI: 10.1016/j.acn.2006.05.007.

[8] Gualtieri, C. and Johnson, L., "Medications do not necessarily normalize cognition in ADHD patients," *Journal of Attention Disorders*, vol. 11(4), pp. 459-469, 2008. DOI: 10.1177/1087054707305314.

[9] Gualtieri, C. and Johnson, L., "A computerized test battery sensitive to mild and severe brain injury," *Medscape Journal of Medicine*, vol. 10(4), pp. 90, 2008.

[10] Gualtieri, C. and Johnson, L., "Neurocognitive testing supports a broader concept of mild cognitive impairment," *American Journal of Alzheimer's Disease and Other Dementias*, vol. 20(6), pp. 359-366, 2005.

[11] Iverson, G. L., Brooks, B. L., Langenecker, S. A., and Young, A. H., "Identifying a cognitive impairment subgroup in adults with mood disorders," *Journal of affective disorders*, vol. 132(3), pp. 360-367, 2011. DOI: 10.1016/j.jad.2011.03.001.

[12] Brooks, B. L. and Barlow, K. M., "A methodology for assessing treatment response in Hashimoto's encephalopathy: a case study demonstrating repeated computerized neuropsychological testing," *Journal of child neurology*, vol. 26(6), pp. 786-791, 2011. DOI: 10.1177/0883073810391532.

[13] A. Rey. L'examen clinique en psychologie. Paris: Presses Universitaires de France, 1964.

[14] Taylor, E.M., *The appraisal of children with cerebral deficits*, Cambridge, MA: Harvard University Press, 1959.

[15] Smith, A., Symbol Digit Modalities Test (SDMT). Manual (Revised). Los Angeles: Western Psychological Services, 1982.

[16] Stroop, J., "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, pp. 643-662, 1935.

[17] Le, T. H., Pardo, J. V., and Hu, X., "4 T-fMRI study of nonspatial shifting of selective attention: cerebellar and parietal contributions," *Journal of Neurophysiology*, vol. 79(3), pp. 1535-1548, 1998.

[18] Nagahama, Y., Sadato, N., Yamauchi, H., Katsumi, Y., Hayashi, T., Fukuyama, H., and Yonekura, Y., "Neural activity during attention shifts between object features," *Neuroreport*, vol. 9(11), pp. 2633-2638, 1998.

[19] Rosvold, H. E. and Delgado, J. M., "The effect on delayed-alternation test performance of stimulating or destroying electrically structures within the frontal lobes of the monkey's brain," *Journal of Comparative and Physiological Psychology*, vol. 49(4), pp. 365-372, 1956.

[20] Fabrigar, L. R., Wegener, D. T., MacCallum, and Strahan, E. J., "Evaluating the use of exploratory factor analysis in psychological research," *Psychological methods*, vol. 4(3), pp. 272-299, 1999.

[21] Kail, R. and Salthouse, T. A., "Processing speed as a mental capacity," *Acta Psychologica*, vol. 86(2-3), pp. 199-225, 1994. DOI: 10.1016/0001-6918(94)90003-5.

[22] Salthouse, T. A., "The processing-speed theory of adult age differences in cognition," *Psychological review*, vol. 103(3), pp. 403-427, 1996.

[23] Stough, C., Thompson, J. C., Bates, T. C., and Nathan, P. J., "Examining neurochemical determinants of inspection time: development of a biological model," *Intelligence*, vol. 29(6), pp. 511-522, 2001.

[24] Grudnik, J. L. and Kranzler, J. H., "Meta-analysis of the relationship between intelligence and inspection time," *Intelligence*, vol. 29(6), pp. 523-535, 2001. DOI: 10.1016/S0160-2896(01)00078-2.

[25] Kranzler, J. H. and Jensen, A. R., "Inspection time and intelligence: A meta-analysis," *Intelligence*, vol. 13(4), 329-347, 1989. DOI: 10.1016/S0160-2896(89)80006-6.

[26] Luciano, M., Wright, M. J., Smith, G. A., Geffen, G. M., Geffen, L. B., and Martin, N. G., "Genetic Covariance Among Measures of Information Processing Speed, Working Memory, and IQ," *Behavior Genetics*, vol. 31(6), pp. 581-592, 2001. DOI: 10.1023/A:1013397428612.

[27] Neubauer, A. C. and Bucik, V., "The mental speed—IQ relationship: unitary or modular?," *Intelligence*, vol. 22(1), pp. 23-48, 1996. DOI: 10.1016/S0160-2896(96)90019-7.

[28] Neubauer, A. C. and Knorr, E., "Elementary cognitive processes in choice reaction time tasks and their correlations with intelligence," *Personality and Individual Differences*, vol. 23(5), pp. 715-728, 1997. DOI: 10.1016/S0191-8869(97)00108-6.

[29] Rindermann, H. and Neubauer, A., "Processing speed, intelligence, creativity, and school performance: Testing of causal hypotheses using structural equation models," *Intelligence*, vol. 32(6), pp. 573-589, 2004. DOI: 10.1016/j.intell.2004.06.005.

[30] Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., and Rabbitt, P., "Cambridge Neuropsychological Test Automated Battery (CANTAB): A Factor Analytic Study of a Large Sample of Normal Elderly Volunteers," *Dementia and Geriatric Cognitive Disorders*, vol. 5(5), pp. 266-281, 1994. DOI: 10.1159/000106735.

[31] Smith, P. J., Need, A. C., Cirulli, E. T., Chiba-Falek, O., and Attix, D. K., "A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with "traditional" neuropsychological testing instruments," *Journal of clinical and experimental neuropsychology*, vol. 35(3), pp. 319-328, 2013. DOI: 10.1080/13803395.2013.771618.

[32] Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Barbas, N., Heidebrink, and J. Giordani, B., "Validity of a brief computerized cognitive screening test in dementia," *Journal of geriatric psychiatry and neurology*, vol. 25(2), pp. 89-99, 2012. DOI: 10.1177/0891988712447894.

[33] Overton, E. T., Kauwe, J. S. K., Paul, R., Tashima, K., Tate, D. F., Patel, and P. Clifford, D. B., "Performances on the CogState and standard neuropsychological batteries among HIV patients without dementia," *AIDS and behavior*, vol. 15(8), pp. 1902-1909, 2011. DOI: 10.1007/s10461-011-0033-9.

[34] Maerlender, A., Flashman, L., Kessler, A., Kumbhani, S., Greenwald, R., Tosteson, T., and McAllister, T., "Examination of the construct validity of ImPACTTM computerized test, traditional, and experimental neuropsychological measures," *The Clinical neuropsychologist*, vol. 24(8), pp. 1309-1325, 2010. DOI: 10.1080/13854046.2010.516072.

[35] Elwood, R. W., "MicroCog: assessment of cognitive functioning," *Neuropsychology review*, vol. 11(2), pp. 89-100, 2001.

[36] Gualtieri, C. and Morgan, D., "The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an

unaccounted source of variance in clinical trials," *The Journal of Clinical Psychiatry*, vol. 69(7), pp. 1122-1130, 2008.

[37] Gualtieri, C. T. and Johnson, L. G., "ADHD: Is Objective Diagnosis Possible?," *Psychiatry (Edgmont (Pa.: Township))*, vol. 2(11), pp. 44-53, 2005.

[38] Meehan, W. P., 3rd, d' Hemecourt, P., Collins, C. L., Taylor, A. M., and Comstock, R. D., "Computerized neurocognitive testing for the management of sport-related concussions," *Pediatrics*, vol. 129(1), pp. 38-44, 2012. DOI: 10.1542/peds.2011-1972.

[39] Register-Mihalik, J. K., Kontos, D. L., Guskiewicz, K. M., Mihalik, J. P., Conder, R., and Shields, E. W., "Age-related differences and reliability on computerized and paper-and-pencil neurocognitive assessment batteries," *Journal of athletic training*, vol. 47(3), pp. 297-305, 2012. DOI: 10.4085/1062-6050-47.3.13.

[40] Ivins, B. J., Kane, R., and Schwab, K. A., "Performance on the Automated Neuropsychological Assessment Metrics in a nonclinical sample of soldiers screened for mild TBI after returning from Iraq and Afghanistan: a descriptive analysis," *The Journal of head trauma rehabilitation*, vol. 24(1), pp. 24-31, 2009. DOI: 10.1097/HTR.0b013e3181957042.