A Computerized Neurocognitive Test to Detect Malingering

A Study of True and Directed Malingerers

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

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

***Abstract*-In order for a cognitive test to be reliable and clinically useful, it ought to have validity parameters that will detect subjects who are exaggerating their deficits. This is the first study of which we are aware to systematically evaluate the problem of invalid response patterns in a computerized test battery (CNT) that is widely used in research and clinical practice. We conclude that validity indicators embedded in the CNT computerized test battery can identify "non-credible responders", i.e., malingerers. Used in conjunction with other tests, the CNT test battery is a quick and efficient way to identify patients who may be malingering cognitive dysfunction. The validity indicators can also be used to exclude invalid test data when the test battery is used in research.**

Keywords- Computerized Test; Malingering; Non-credible Responding; Reliability

I. INTRODUCTION

As computerized neurocognitive tests (CNTs) are used more often as stand-alone screening tests or as complements to a conventional neuropsychological battery, it is appropriate to examine methods for determining whether the results they generate are valid. This point was emphasized in a joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology [1]. When CNTs are used, for example, in clinical trials, it is essential to exclude invalid data. When athletes are pre-screened with a "concussion management battery," it is important to know if the subject is giving less than maximal effort to protect his playing time if he is concussed. Unscrupulous litigants pretend to have suffered brain injuries, and university students seeking stimulant drugs pretend to have attention deficit/hyperactivity disorder (ADHD). A test battery used to detect concussions, brain injuries, or ADHD ought to be able to identify patients who are not making a sincere effort, in order to manipulate the outcome.

Invalid response patterns may reflect failure to understand the test or comply with the instructions. Transient confusion, cognitive impairment, or resistance to the assessment process may lead to invalid results. Alternatively, an invalid response pattern may indicate willful exaggeration or manipulation, i.e., malingering. This study is concerned with the latter case, but it is likely that the same indicators developed in the study of malingering are relevant to other areas where test validity is an issue.

The neuropsychological assessment of malingered cognitive deficits traditionally relies on tests designed for that specific purpose, e.g., the Victoria Symptom Validity Test [2], the Rey 15 item test [3, 4], the Computerized Assessment of Response Bias (CARB) [5], the Test of Memory Malingering (TOMM) [6], and others. An alternative method is to examine the results of standard neuropsychological tests to detect an error pattern that is excessive for or incompatible with known psychiatric or neurological disorders. In the latter case, validity indicators are "embedded" in the fabric of a test that is ordinarily used for memory assessment, attention, executive function, etc. The advantages are efficiency – one test serving two purposes – as well as protection from coaching to prepare for well-known validity tests. Another advantage is that aggregating multiple independent validity indicators can improve diagnostic accuracy [7]. Furthermore, embedded measures of effort throughout a test battery allow for a continuous assessment of one's effort level, which may vary across the test battery. Embedded measures are also immune to the argument that effort test failure does not equate to failure on other tests within a battery, because the failure occurs within the clinical test in question [8].

Whether or not a test is used to uncovering malingerers, it is necessary to know how to assay the validity of the test results. The method we used in this report was naturalistic: a large number of patients who had severe traumatic brain injuries were compared to a group of suspected malingerers and to another group of people who had been instructed to behave as if they were malingering a brain injury.

II. METHOD

A. The Clinical Database

When a patient is evaluated at one of the North Carolina Neuropsychiatry Clinics (in Chapel Hill, Charlotte or Raleigh), he or she is routinely administered a computerized neurocognitive test battery (CNT), along with rating scales, psychological tests, and validity measures, as appropriate. The CNT data is automatically uploaded to a central database, which is maintained under secure conditions and available only to selected clinicians in the practice. Neuropsychiatric diagnoses are based on a comprehensive examination, of which the CNT battery is only a part; diagnoses are not made simply on the basis of CNT scores. For this study, diagnoses were affirmed by review by a research psychiatrist (Gualtieri). Patients give written informed consent to the use of their de-identified clinical data, including the CNT, for research purposes and program evaluation.

B. Participants

In the database maintained by the Neuropsychiatry Clinics, there are data from more than 13,000 neuropsychiatric patients and 4400 normal controls who took the CNT test battery. In this database, we identified four groups of 20 to 65 years of age: 2172 normal participants (NML), 589 patients with traumatic brain injury (TBI), and a group of 40 people who were evaluated in the clinic and strongly suspected of malingering (SM). A conservative threshold was used to make the determination that a subject was likely to be malingering. This may explain the relatively low number comprising the SM group relative to the TBI group (6.4%). It should be noted, however, that this conservative position increases the possibility that some of the individuals in the TBI group should have been in the SM group. To the clinic database we added a group of 60 "directed malingerers" (DMAL), normal people who were instructed to behave as if they had a disabling brain injury.

The normal control participants were in good health and had no history of a psychiatric or neurological disorder; they were taking no centrally active medication. They were recruited from the community when norms were developed for the CNT, and their data was de-identified. They signed written informed consent for use of their data. Their demographic characteristics matched those of the latest (2010) US census. TBI patients all had had a severe injury, with a Glasgow Coma Scale score < 8 when seen in the emergency department. They were more than one year past the injury and had recovered sufficiently to take the CNT, which requires bimanual dexterity, good vision, and a fourth grade reading ability. Demographic variables are presented in Table 1. Although mean values were similar between groups in most respects, statistically significant differences were identified. The implications of these differences are discussed where relevant.

TABLE 1 DEMOGRAPHIC CHARACTERISTICS OF THE SAMPLES

	NML		TBI		SM		DMAL		ANOVA Sig <	EFF SIZE cp to NML*		
N	2172	sd	589	sd	40	sd	60	sd		TBI	SM	DMAL
AGE	41.1	12.2	42.3	11.7	44.1	9.9	38.9	12.7	0.029	0.10	0.25	0.19
COMPNUM	2.79	0.46	2.30	0.66	1.83	0.80	2.88	0.33	0.000	0.00	0.67	0.10
EDUC	16.2	2.4	13.8	2.8	14.0	3.8	16.2	2.2	0.000	0.88	1.72	0.16
% MALE	0.37		0.70		0.83		0.50		0.000			
% WHITE	0.87		0.89		0.63		0.77		0.000			

**The effect size by Cohen's d comparing each group to normal Ss.*

Forty participants were included in the SM group. These were patients whom we believed were trying to prove they were cognitively disabled, even though they really were not. The CNT was used in their evaluation. However, the measures of validity were considered after the individuals were evaluated and did not play a role in group selection for the present study. Many, if not most, of the suspected malingerers have been followed up in our clinic or through medical case managers; in those cases, the designation of malingering was supported by subsequent events. Each SM exhibited the following characteristics:

1. Their symptoms were notably out of proportion to (a) the nature of the injury, (b) known pathology, and (c) objective signs of actual disease.
2. The patient had something tangible to gain by virtue of illness, pain or disability, establishing secondary gain.
3. The patient did not have an active psychiatric disorder (e.g., depression) or neurological disorder (e.g., epilepsy). He or she may, however, have had a personality disorder, substance abuse or alcoholism, or a criminal history.
4. The patient was administered the Rey 15 item test, as well as one established performance validity test: the TOMM, the Victoria Symptom Validity Test, or the CARB. They failed at least one of those measures.
5. The patient was observed, during the evaluation, to do things that he claimed he could not do, or behaved in a manner that was incompatible with his claimed disability (e.g., an excellent memory for interactions with physicians and lawyers, but with memory testing in the dementia range; complaints of extreme pain, fatigue, etc., but the "patient" was able to comply with a long and arduous examination with no evident distress.)

In addition, all of the patients exhibited at least two of the following four characteristics:

1. The patient had non-physiological physical signs (e.g., monocular diplopia, glove or stocking hypesthesia, Waddell's signs).
2. The patient had non-physiological mental symptoms or signs: e.g., dense retrograde amnesia, absent anterograde amnesia, or atypical/improbable auditory hallucinations.
3. Other physicians or psychologists had suspected non-credible illness behavior.
4. There was external evidence (e.g., surveillance) that indicated the patient was capable of doing what he claimed he could not do.

The "directed malingers" (DMAL) were normal individuals who were recruited to take the CNT test while playing a specific role. They were to take the test as if they had had a head injury from a motor vehicle accident and wanted to convince the doctor that they had persistent and disabling cognitive problems. The participants were nonprofessionals and unfamiliar with clinical protocols used to detect malingering. Before they took the CNT, they read a script that has been previously used in published research of feigned malingering [9] (Appendix 1).

C. The CNT Battery

The CNT used in this study, an updated version of the CNS Vital Signs test battery, is a screening measure of cognitive functioning that contains seven tests and generates eight test scores. The CNS Vital Signs test battery has been widely used by neurologists, psychiatrists, and neuropsychologists [10]. Seven of the 8 test scores in the CNT load onto three factors: memory, attention, and information processing speed. The tests generate raw scores and standard scores. Scores are standardized by adjusting for age and education, based on data from 3,420 normal participants (age 4 to 90). The CNT battery generates both raw scores and standard scores. Standard scores are reported with a mean of 100 with a standard deviation of 15. A single summary score, the Index score (INDEX), is computed by averaging the standard scores of the three factor scores (Table 2). Factor scores and the Index are derived from standard scores.

TABLE 2 THE CNT

TEST	ABBREV	FACTOR	SCORE
Verbal Memory Test	VBM	MEMORY	Correct responses minus errors
Visual Memory Test	VIM	MEMORY	Correct responses minus errors
Finger Tapping Test	FTT	*	Total number of taps, right and left
Symbol Digit Coding Test	SDC	PROCESSING SPEED	Correct responses minus errors in two minutes
Stroop Test	RT	PROCESING SPEED	Average of complex and Stroop response times
	ST	ATTENTION	Number of errors in non-congruent condition
Shifting Attention Test	SAT	PROCESSING SPEED	Correct responses minus errors
Continuous Performance Test	CPT	ATTENTION	Correct responses minus errors

**The finger tapping test does not load onto to any of the three factors.*

Test-retest reliability and the concurrent validity of CNT are comparable to similar, conventional neuropsychological tests [10]. Research has indicated relevance to the study of mood disorders [11, 12], sleep disorders [10, 13] encephalopathy [11, 14] post-concussion syndrome, and severe traumatic brain injury [15].

D. Analysis

This study utilized a known groups design. The scores of the four groups on each of continuous variables of the CNT were analyzed by: multiple analysis of variance (MANOVA), controlling for age, race, gender, education, and computer familiarity, because the four groups differed significantly in all of those variables; one-way analysis of variance with Bonferroni correction or MANOVA with Tukey's HSD; and pairwise t tests, as specified. Sensitivity and specificity analysis was performed by generating a receiver operating characteristics (ROC) curve (Statistical Package for the Social Sciences, SPSS 17).

III. RESULTS

A. Comparing Means

Fig. 1 presents the mean standard scores for each of the four study groups across eight test scores. Standard scores are presented so the data from different measures are comparable. All eight measures showed the same pattern: normal participants performed best and TBI patients performed second best. The SM and DMAL groups scored lowest in all of the measures. All of the group differences were significant at the $P < 0.0001$ level (MANOVA).

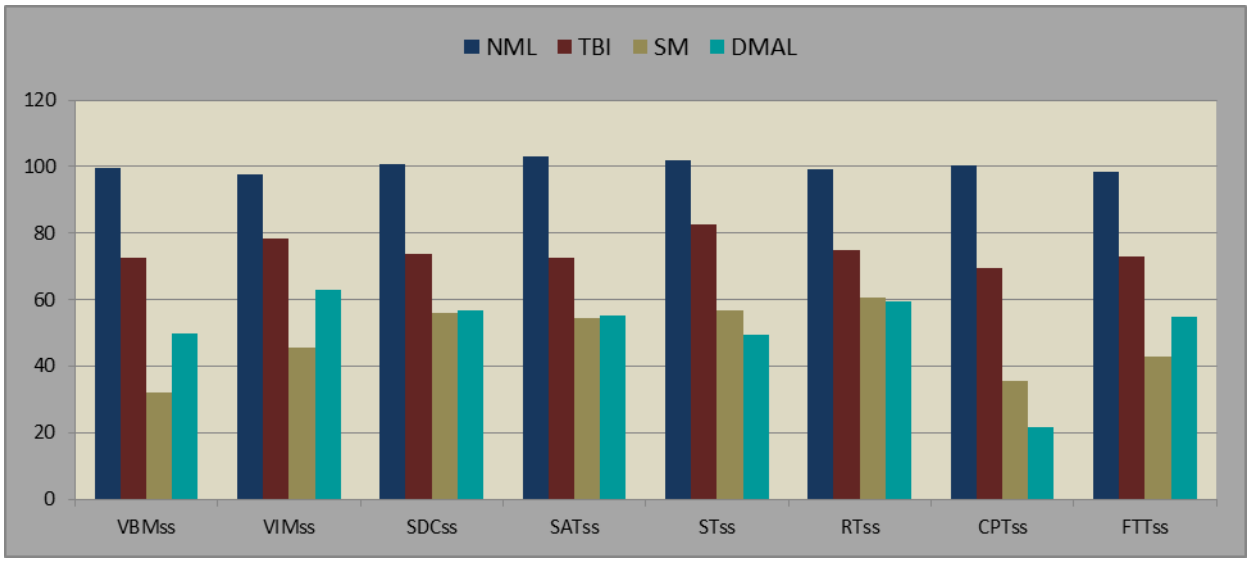

Fig. 1 Cognitive performance of the four groups

Note, Fig. 1: the y axis is test standard score. NML, normal subjects; TBI, patients with brain injuries; SM, suspected malingerers; DMAL, directed malingerers; VBMss, verbal memory test standard score; VIM, visual memory test; SDC, symbol digit coding test; SAT, shifting attention test; ST, Stroop test; RT, Stroop test reaction time; CPT, continuous performance test; FTT, finger tapping test. All of the group differences were significant at the P < 0.0001 level (MANOVA).

When the three clinical groups were compared to normals, effect size (Cohen's d) ranged from 0.81 to 2.84. When the four groups were compared by one-way analysis of variance with Bonferroni correction, the differences between the four groups were statistically significant at the P < 0.01 level, with these exceptions: the SM and DMAL groups did not differ on the SDC, SAT, ST, RT, or CPT. When SM and DMAL were compared by MANOVA, controlling for age, race, gender, education, and computer familiarity, none of the test scores were statistically significant.

B. Validity Indicators

Having established that the SM and DMAL groups scored significantly lower than the TBI and NML groups and that their scores did not differ significantly from one another, we were assured that the groups were appropriate for further investigation. The following three validity indicator studies examined the salience of different but overlapping parameters for detecting non-credible response patterns.

Three approaches to validity detection were evaluated for their sensitivity and specificity: the Index score, the eight standardized test scores, and the test raw scores, including reaction times. Because these three approaches are non-independent, the probabilities of invalid responding cannot be chained. However, probabilities derived from the raw scores can be because they are independent measures.

ROC (receiver operating characteristic) analysis generates two elements: the sensitivity and specificity of a test. Sensitivity speaks to the likelihood that a true positive, in this case, an individual in the SM group, will be correctly identified. Specificity is the proportion of true negatives that are correctly identified, i.e., patients with TBI. ROC analysis was performed by comparing TBI patients to the SM group and to the DMAL group; we consider the first comparison to be more pertinent to clinical practice (Table 3). On the basis of the TBI/SM comparison, specificity curves can be generated for each test, indicating the likelihood that someone who scores above a certain score is a member of the SM group. Conversely, the variable 1 - specificity gives the probability that the test has not misclassified a TBI patient in the SM group. A specificity curve can be drawn for all of the relevant variables indicating the likelihood of an invalid response (Figs. 2 and 3).

TABLE 3 SENSITIVITY AND SPECIFICITY, TBI VS. SM AND TBI VS. DMAL

	SUSPECTED MALINGERERS			DIRECTED MALINGERERS		
	Area under curve	Sensitivity	Specificity	Area under curve	Sensitivity	Specificity
INDEX 10 < 45	0.816	0.436	0.896	0.760	0.367	0.883
6 OR MORE STANDARD SCORES < 70	0.806	0.550	0.837	0.756	0.550	0.837
3 OR MORE TEST VALIDITY INDICATORS	0.886	0.632	0.884	0.797	0.524	0.85
VBM raw score < 34	0.869	0.688	0.883	0.717	.356	0.914

VIM raw score < 34	0.837	0.676	0.797	0.676	.373	0.845
FTT total below 40	0.811	0.514	0.912	0.682	.259	0.914
SDC correct ≤ 20	0.746	0.474	0.852	0.665	.458	0.852
Negative Stroop effect	0.639	0.469	0.839	0.716	.583	0.841
SAT errors ≥ SAT correct	0.719	0.571	0.742	0.656	.383	0.731
Simple RT ≥ choice RT	0.795	0.771	0.749	0.669	.682	0.753
CPT correct < 30	0.745	.405	0.881	0.788	.500	0.881
AVERAGE	**0.788**	**0.562**	**0.843**	**0.717**	**0.458**	**0.846**

Note, Table 3. Sensitivity and specificity by ROC. VBM, verbal memory test; VIM, visual memory test; SDC, symbol digit coding test; SAT, shifting attention test; ST, Stroop test; RT, reaction time; CPT, continuous performance test; FTT, finger tapping test.

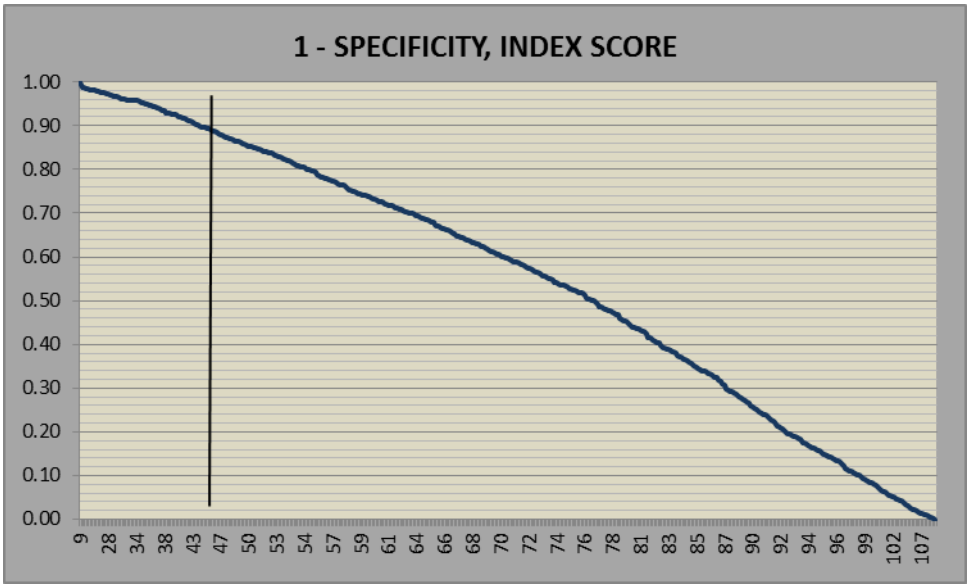

Fig. 2 Specificity curve, index score

Note, Fig. 2: x axis, Index score; y axis, probability of noncredible responding; ROC analysis.

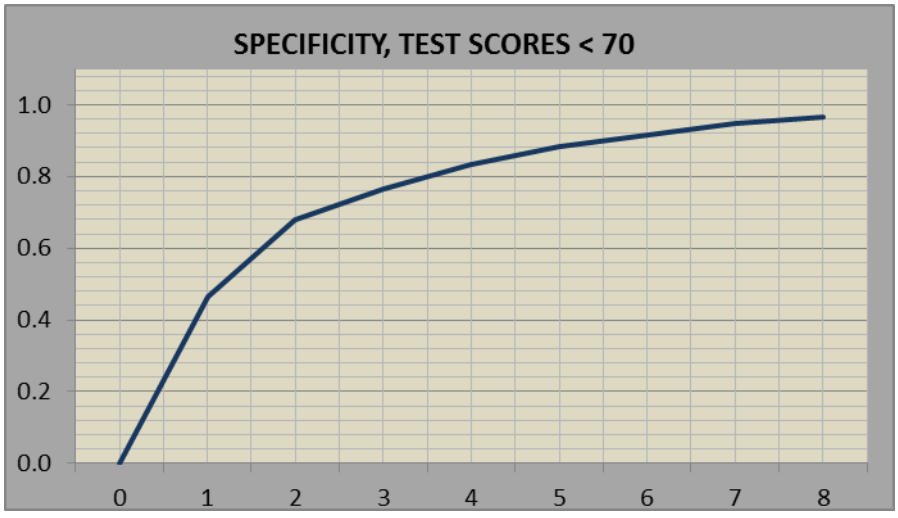

Fig. 3 The specificity curve for test70vi

Note, Fig. 3: x axis, number of test scores < 70; y axis, probability of noncredible responding; ROC analysis.

In Table 3, we list twelve measures that were the most sensitive and specific for the detection of malingerers. In the same table, we give the area under the curve (AUC), specificity, and sensitivity data for the TBI vs SM and TBI vs DMAL comparisons.

1) Validity Indicator 1: Index score less than 45

Fewer than one in ten thousand normal individuals will generate an Index score of 45. In the NCNP database, which includes patients with early dementia and mild mental retardation, less than 1.8% scored lower than 45. In this sample of

NMLs, none scored lower than 45, and only 10% of TBI patients did (Table 4).

TABLE 4 LOW INDEX SCORES IN THE 4 GROUPS

	INDEX < 45	INDEX < 37
NML	0.00	0.00
TBI	0.10	0.05
SM	0.45	0.39
DMAL	0.35	0.15

Note, Table 4. NML, normal subjects; TBI, patients with brain injuries; SM, suspected malingerers; DMAL, directed malingerers; INDEX, index score.

The statistic of interest here is specificity: the proportion of actual negatives (i.e., non-malingerers) that are correctly identified. The inverse (1 – specificity) indicates the false positive rate, that is, the likelihood that a true negative will not be identified as a malingerer. Fig. 2 indicates that an Index score of 45 yields an inverse specificity score of 0.90; that is, if a participant scores less than 45, there is a 90% probability that he or she is faking. If less than 37, there is a 95% probability.

2) Validity Indicator 2: Number of standard scores below 70 (Test70VI)

The CNT generates eight scores, and these are reported as raw scores and standard scores. A standard score of 70 indicates that the subject scored 2 or more standard deviations below the population mean. A second approach to validity assessment, therefore, might be the number of test scores in the CNT that are less than 70 (TEST70VI) (Table 5). When this parameter is applied to the comparison of TBI patients to SM, the ROC area under the curve is 0.806. The criterion of 90% is achieved if 6 scores are below 70. The specificity curve for TEST70VI (TBI v MAL) is given in Fig. 3. From this graph, it is possible to calculate the probability of incorrect identification relative to the number of tests scores below 70 (Table 6).

TABLE 5 OCCURRENCE OF TEST SCORES < 70 IN THE FOUR GROUPS

	TEST70>3	TEST70>4	TEST70>5
NML	0.001	0	0
TBI	0.27	0.19	0.1
SM	0.65	0.65	0.55
DMAL	0.73	0.65	0.47

Note, Table 5. NML, normal subjects; TBI, patients with brain injuries; SM, suspected malingerers; DMAL, directed malingerers; INDEX, index score. TEST70 is number of test standard scores above 70.

TABLE 6 PROBABILITY OF NON-CREDIBLE RESPONSE PATTERN BY NUMBER OF TEST SCORES LESS THAN 70

Tests < 70	Probability
0	0%
1	46%
2	68%
3	76%
4	83%
5	88%
6	92%
7	95%
8	97%

3) Validity Indicator 3: TEST VI's based on raw scores

The test raw scores were screened by ROC analysis. Tests were chosen for evaluation on the basis of comparison between TBI patients and the SM group if the area under the ROC curve (AUC) was greater than 0.7. The exception was a negative Stroop effect, because it is an intuitive indicator of invalid responding if the participant's reaction time is faster in the incongruent condition than the congruent. The eight test VI's are listed in Tables 3 and 7. In contrast to the two previous VI's, these occur only occasionally in normal subjects and in one of seven TBI patients.

TABLE 7 THE EIGHT TEST VI'S IN FOUR GROUPS; % OCCURRENCE

	NML	TBI	SM	DMAL
VBM raw score < 34	0	0.08	0.6	0.35
VIM raw score < 34	0	0.14	0.65	0.38
FTT total below 40	0.003	0.08	0.45	0.25
SDC correct ≤ 20	0.02	0.14	0.45	0.45
Negative Stroop effect	0.08	0.14	0.38	0.58
SAT errors ≥ SAT correct	0.04	0.21	0.68	0.45
Simple RT ≥ choice RT	0.01	0.24	0.53	0.38
CPT correct < 30	0.02	0.11	0.38	0.37
Average	0.02	0.14	0.52	0.40

Note, Table 7. VBM, verbal memory test; VIM, visual memory test; SDC, symbol digit coding test; SAT, shifting attention test; ST, Stroop test; RT, reaction time; CPT, continuous performance test; FTT, finger tapping test.

4) "Chaining" Test Results

The sensitivity of the tests in Table 3 is low, averaging 0.568 for the SM comparison and 0.461 for the DMAL comparison. This implies that no one test is sufficient as a stand-alone test to detect malingering. However, since the eight TEST VI's are independent tests—that is, they can be administered in standalone fashion just as any validity test—the probabilities generated by ROC analysis can be "chained" to generate a conditional probability, relative to the number of tests that indicate invalid responding [12]. If p + q = 1, where the probability of a true negative (specificity) is q, then the probability of a false positive is p, or 1-q. The number of validity indicators is n and the number of invalidity indicators is n1. The number of tests that do not meet criterion is n-n1. The formula, as per Larrabee & Berry, is:

$$\text{Probability of a positive result by chance alone} = ((n!/(n1!*(n-n1)!)*(p^{n}*q^{n-1}).$$

In Table 8, the probability of misidentifying a real patient as an invalid responder decreases dramatically as more of the tests are invalid. A participant who scores below criterion on three tests has a 20% chance of being incorrectly identified as an invalid responder. With four tests, 9%; with five, 3%, and so on. No NML subjects had more than two test VI's. Only 5% of the TBI patients had more than three test VI's, but 42% of the SM group did, and 52% of the DMAL group.

TABLE 8 PROBABILITY OF A FALSE POSITIVE IDENTIFICATION

# INVALID TESTS	PROBABILITY
1	0.269
2	0.296
3	0.198
4	0.089
5	0.029
6	0.007
7	0.001
8	0.000
9	0.000
10	0.000
11	0.000
12	0.000

5) Validity Indicator 4: The two memory tests are forced-choice tests

The verbal and visual memory tests in CNT are recognition memory tests. In each test, there are 60 items and to each item the participant is asked to respond "yes" if the item is correct or "no" if the item is incorrect. The VBM and VIM also happen to be forced choice tests; that is, pressing the button every time, or not pressing the button at all, should generate a test score of 30 on VBM or VIM. Theoretically, a score less than 30 on VBM or VM indicates willful exaggeration. Someone who simply presses the button randomly, or does not press the button at all, ought to score 30. Therefore, a score less than 30 is not only an invalid response, but one suggestive of willful manipulation.

None of the 2172 normal participants scored less than 30 on VBM or VIM. Among 589 TBI patients, 14 (2%) scored below 30 on VBM (n=6) or VIM (n=8). In contrast, among the 40 individuals in the SM group, 17 scored below 30 on VBM (8) or VIM (11), and two scored below 30 on both. In the DMAL group (n = 60), 17 scored below 30 on VBM (12) or VIM (14) and 9 scored below 30 on both.

In the SM group, 5 additional participants were identified by virtue of VBM or VIM scores below 30 who did not meet criterion on any of the other VI's. Applying the two strategies in tandem, therefore, identified 26 out of 40, or 65%. In the DMAL group, an additional 4 participants were identified; if this is added to the 14 who were identified by the VI's, 18 of 60 (30%) were correctly identified.

IV. CONCLUSIONS

A test that is used in research or clinical practice ought to have validity parameters. This is the first study of which we are aware to systematically evaluate the problem of invalid response patterns in a computerized neurocognitive test battery. By comparing the results of suspected malingerers and directed malingerers to patients with severe brain injuries, one is able to develop criteria for determining whether a test is valid or not. These results are relevant to the clinical interpretation of a test.

Appreciating the validity of a participant's performance is also important because tests like CNT are sometimes used to evaluate people who may be pretending to be disabled. By comparing the performance of TBI patients to groups of suspected malingerers and directed malingerers, we were able to extract a number of parameters, or validity indicators, to differentiate the former from the latter with sensitivity and specificity characteristics similar to those of conventional psychological tests [16]. In addition, using the TEST VI parameters allows for the aggregation of multiple validity indicators via the chaining of likelihood ratios findings, which increases discrimination between groups considerably [12] and represents a compelling advantage in comparison to many other measures of test validity. If 4 or more of those VI's are present, it is highly likely (>90%) that the participant's performance was not a valid reflection of his or her cognitive abilities. We recommend using a cutoff of 5 or more TEST VI's, which produces a 97% likelihood that the participant's performance was not a valid reflection

of his or her cognitive abilities. This is a conservative approach to identifying poor effort that may represent malingering when occurring in the proper context in accordance with current standards of practice, e.g., the Slick et al. criteria of malingered neurocognitive disorder [7].

Since two of the tests, VBM and VIM, are forced-choice tests, it is possible to detect what may be willful manipulation by an insincere participant. Forty-three percent of the clinically identified malingerers scored below chance on either VBM or VIM or both, and 28% of the directed malingerers. Only 2% of the TBI patients generated scores below chance on these tests. That may be because they were in fact malingerers whom the clinicians failed to identify. On the other hand, if a participant is only responding at random, by virtue of confusion, fatigue or pain, for example, he may score below chance on one of the tests. In the case of random responding, a score of 29 will occur 9.9% of the time; a score of 28, 8.9%; 27, 7.6%; 26, 6.1%; and 25, 4.5% of the time. Random responding, of course, is clearly indicative of invalid test-taking but not necessarily of wilful exaggeration.

Like other psychological tests that are used to detect invalid responders, the sensitivity of the VI's described here is relatively low. A single test is seldom likely to capture more than 40% of the invalid responders who take it. Since each of the eight TEST VI's is an independent test, one can chain the probability of invalid responding in multiple tests and arrive at a more secure conclusion about the participant who has taken the test. Chaining TEST VI's decreases the probability of misidentifying a valid responder. However, when one applies two criteria, more than five TEST VI's and VBM and/or VIM scores below 30, only 65% of the SM group were successfully identified.

The various VI's in the CNT battery are embedded in the fabric of the test, which is a distinct advantage. However, the low sensitivity of the test is a caution against using the CNT in isolation. The neuropsychiatric or neuropsychological evaluation of possible malingering should always employ more than one test of non-credible responding, and CNT, even with numerous VI's and a forced-choice component, is just one test. It is recommended practice to require more than one invalid test, as well as the clinical criteria set forth earlier, before one can aver that someone is trying to manipulate the results of an evaluation [8, 12].

There are limitations to this study. The four groups were chosen from a convenience sample of data generated in the course of clinical operations. The four groups were disparate in size and important demographic characteristics; the former may lead to potential problems, e.g., unequal variance. The essential comparator, TBI patients, was a reasonable choice, because all of the SM participants were trying to convey cognitive impairment. However, one variable that was not controlled was the amount of time that passed following injury before the TBI patient was assessed. Other comparison groups might also have been interesting: for example, patients with conversion disorder or somatoform disorders, chronic pain, chronic mental illness, or post-traumatic stress disorder. These conditions are sometimes relevant to the differential diagnosis of a patient who is presenting himself as disabled. In fact, in earlier analyses not described herein, we found that patients with those conditions performed similarly on CNT to TBI patients, and score much better than the SM and DMAL participants. The SM and DMAL participants, as a group, score lower on CNT than patients with early dementia or mild mental retardation.

We are confident that the suspected malingerers we labelled as such were properly identified. The Slick criteria [7] were not applied in a prospective fashion, but the criteria used to define this group are consistent with those criteria. A consistent series of validity tests was not used across all patients, but in all cases at least two measures were used and one was failed. Before tests were administered, it was clear that we were dealing with suspicious cases: litigation over disability was at issue, patients who appeared to be in robust good health were complaining of extraordinary symptoms relative to minor injury, patients who said they could hardly move without pain and spent their days on the couch had grease-stained hands, and patients who could recount their ill-treatment by employers and insurance companies in outraged detail scored in the dementia range on memory tests. In some cases, the very fact that they were able to complete the CNT protocol is suspicious, when they perform worse on CNT than patients with early dementia or mild mental retardation, who usually require assistance to take the test battery.

The strict criteria we applied to identify malingerers probably consigned some fakers to the TBI group. This represents a weakness, to be sure, but a necessary weakness in the clinical arena, where the identification of a patient as a malingerer should be done with caution, and where it is proper to give every patient the benefit of the doubt. In the event that some true malingerers were misidentified as TBI patients, this misidentification would only weaken the statistical comparisons and conspire against our hypotheses.

In the neuropsychological literature, recourse is often made to "feigned" or "directed" malingerers; normal people who are instructed to behave as if they had had a disabling brain injury [9]. The rationale is that one can be quite sure, in such an event, that the participant really was "faking bad," as opposed to the real world, where there is always an element of uncertainty. The only way you can be sure that someone is really lying, after all, is if he or she admits they were lying. Accepting this line of thought, albeit with reservations, we compared the performance of 60 feigned malingerers to the 40 patients we thought were really malingering. There were small differences between the SM and DMAL groups, but they were not statistically significant when demographic variables were controlled.

Our group of feigned malingerers was not a perfect match for the "real malingerers." Compared to the latter, the directed

malingerers were more likely to be female, white, and better educated (Table 1). They did, however, manage to generate scores that were as low as the scores generated by malingering patients, and dramatically lower than patients with TBI brain injury. More to the point, the profile generated by the DMAL group was quite similar to that of the SM group.

These results indicate that the CNT includes an embedded system for reliably identifying suspected malingering with acceptable specificity of a clinical control group using multiple measures. This study represents a step advancing the development establishing neuropsychological test validity, specifically with respect to computerized assessment, in accord with recent recommendations [1]. This study also represents the first attempt at a comprehensive validity system for a computerized cognitive test battery. If, therefore, it is necessary or desirable to utilize computerized tests in the course of patient evaluation, the CNT battery has VI's to indicate whether a specific test score is invalid. If multiple VI's are present, one is advised to discount the validity of the test battery as a whole. It is the clinician's responsibility to discover why an individual participant performed so poorly on the test battery.

APPENDIX

1. Study instructions

"Imagine that you were in a car accident in which another driver hit your car. You were knocked unconscious, and woke up in the hospital. You were kept overnight for observation. The doctors told you that you experienced a concussion. Try to imagine that a year after the accident, you are involved in a lawsuit against the driver of the other car. If you are found to have experienced significant injuries as a result of the accident, you are likely to receive a bigger settlement. You have decided to fake or exaggerate symptoms of a brain injury in order to increase the settlement you will receive. As a part of the lawsuit, you are required to undergo computerized cognitive testing to determine whether or not you have experienced a brain injury. If you can successfully convince the individual examining the results of the computerized testing that you have experienced significant brain damage, you are likely to get a better settlement. If the examiner detects that you are faking, you are likely to lose the lawsuit. You are about to take a series of cognitive tests that would be used in such a situation. We would like you to simulate brain damage, but in a believable way, such that your examiner cannot tell that you are attempting to fake a brain injury."

REFERENCES

[1] 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.

[2] Slick, D. J., Hopp, G., Strauss, E., and Spellacy, F. J., "Victoria Symptom Validity Test: efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales," *Journal of Clinical and Experimental Neuropsychology*, vol. 18(6), pp. 911-922, 1996.

[3] Reznek, L., "The Rey 15-item memory test for malingering: a meta-analysis," *Brain Injury: BI*, vol. 19(7), pp. 539-543, 2005.

[4] Rey, A., L'Examen Clinique en Psychologie. Paris: Presses Universitaires de France, 1964.

[5] Green, P. and Iverson, G. L., "Validation of the computerized assessment of response bias in litigating patients with head injuries," *The Clinical Neuropsychologist*, vol. 15(4), pp. 492-497, 2001.

[6] Tombaugh, T., "The Test of Memory Malingering," *Toronto: Multi-health systems*, 1996.

[7] Slick, D. J., Sherman, E. M., and Iverson, G. L., "Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research," *The Clinical Neuropsychologist*, vol. 13(4), pp. 545-561, 1999.

[8] Boone K.B., *Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective*, New York: Guilford Press, 481 pages, 2007.

[9] Suhr, J. and Boyer, D., "Use of the Wisconsin Card Sorting Test in the Detection of Malingering in Student Simulator and Patient Samples," *Journal of Clinical and Experimental Neuropsychology*, vol. 5(21), pp. 701-708, 1999.

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

[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] Larrabee, G. and Berry, D., "Diagnostic classification statistics and diagnostic validity of malingering assessment," *Assessment of Malingered Neuropsychological Deficits*, New York: Oxford University Press, pp. 14-26, 2007.

[13] Armstrong, L., Glidewell, R., Orr, W., and Roby, E., "The Utility Of Using The Apnea-Hypopnea Index And Computer Administered Neuropsychological Testing To Predict CPAP Treatment Adherence: A Retrospective Analysis," *Journal of Sleep and Sleep Disorders Research*, vol. 34, pp. 136, 2011.

[14] 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.

[15] 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.

[16] Vickery, C.D., Berry, D.T.R., Inman, T.H., Harris, M.J., and Orey, S.A., "Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures," *Archives of Clinical Neuropsychology*, vol. 16, pp. 45-73, 2001.