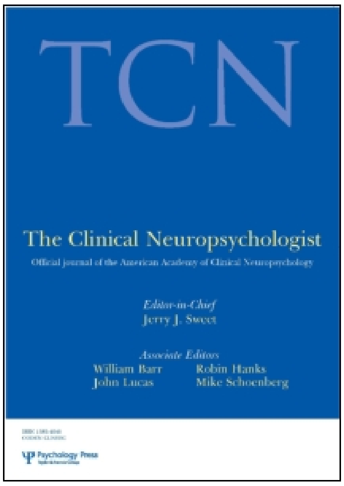

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713721659

Symptom Validity Test Failure Indicates Invalidity of Neuropsychological Tests

David D. Fox[a]
[a] Consultant in Psychological Assessment, Glendale, CA, USA

Online publication date: 11 April 2011

To cite this Article Fox, David D.(2011) 'Symptom Validity Test Failure Indicates Invalidity of Neuropsychological Tests', The Clinical Neuropsychologist, 25: 3, 488 — 495
To link to this Article: DOI: 10.1080/13854046.2011.554443
URL: http://dx.doi.org/10.1080/13854046.2011.554443

The Clinical Neuropsychologist, 2011, 25 (3), 488–495
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854046.2011.554443

Psychology Press
Taylor & Francis Group

Symptom Validity Test Failure Indicates Invalidity of Neuropsychological Tests

David D. Fox
Consultant in Psychological Assessment, Glendale, CA, USA

Neuropsychological and symptom validity test results from 220 archival cases were analyzed to determine if failing a symptom validity test (SVT) affects the relationship between neuropsychological tests and brain damage. Results reveal that among those who failed either the Word Memory Test (WMT) or the Computerized Test of Attention and Memory (CTAM) there was no correlation between the results of 25 commonly used neuropsychological tests and objectively determined brain damage. For those who passed SVTs, the expected relationship between neuropsychological tests and brain damage was found. Consistent with earlier findings, effort had a greater effect on test performance than did brain damage. SVT performance was not correlated with either brain damage or the presence of external incentives. Results indicate the need for symptom validity testing in all cases and that failure of a single SVT can invalidate the expected brain–behavior relationship that underlies neuropsychological test interpretation.

Keywords: Symptom validity testing; SVT; Word Memory Test; Computerized Test of Attention and Memory; Neuropsychological validity.

INTRODUCTION

One of the fundamental tenets of neuropsychology is that there is a relationship between an individual's behavior as measured by neuropsychological tests and the condition of the brain (Lezak, Howieson, & Loring, 2004). However, research in the past two decades has shown that the accuracy or validity of such tests can be compromised by feigning of cognitive deficits or low effort. Specific procedures have been developed to detect malingering or poor effort, often referred to as symptom validity testing (Boone, 2007) and failure of a symptom validity test (SVT) has been shown to indicate an increased probability of malingering (Larrabee, 2007). There is ample inferential evidence that poor performance on an SVT undermines the confidence that can be placed in the results of neuropsychological findings (Green, 2007). In addition the accumulating data on SVTs and the need for such testing have been recognized by professional associations (Bush et al., 2005; Heilbronner et al., 2009). However, in practice, some clinicians downplay the importance of validity measures (e.g., Bigler, 2001) or believe that, even with SVT failure, neuropsychological tests can accurately measure brain-based cognitive impairment. Recently McGrath, Mitchell, Kim, and Hough (2010) reviewed some of the literature on response bias, including SVTs, and described how potential response bias can attenuate or eliminate the criterion-related

Address correspondence to: David D. Fox, 410 Arden Ave Ste 201, Glendale, CA 91203, USA. E-mail: DavidFox@psychotherapy.org
Accepted for publication: January 4, 2011. First published online: February 18, 2011

validity of a substantive indicator. To demonstrate the validity of a bias indicator like an SVT, the correlation of a substantive indicator (e.g., a neuropsychological test) with a criterion (e.g., brain damage) should be greater when the SVT is passed than when it is failed. Based on their review, they concluded that the case remains open whether existing measures have sufficient utility to justify their use.

In line with the issues raised by McGrath et al. (2010), this study was designed to formally measure the relationship between neuropsychological test results and documented brain damage in people who fail as compared to those who pass SVT testing. A previous study found that effort has a greater effect on overall performance on a neuropsychological battery in head-injured patients than does brain damage (Green, Rohling, Lees-Haley, & Allen, 2001). However, that study did not examine whether individual tests might still reflect brain damage even when SVTs are failed. Further, the effect of SVT failure in a general neurological population has not been explored, even those without any apparent incentive. This study addresses the relation of effort to these variables with specific hypotheses:

(1) Those who pass SVTs show the expected relationship between neuropsychological tests and objective measures of neuropathology.
(2) Those who fail SVTs demonstrate no correlation between test performance and objective measures of brain damage.
(3) SVT performance has a greater effect on neuropsychological test results than does brain damage.
(4) SVT performance is not affected by brain damage.

This study examined archival data from 220 individuals who had undergone neuropsychological evaluations, and each person also had completed two specific SVTs, the Word Memory Test (WMT) and the Computerized Test of Attention and Memory (CTAM). These individuals were referred for determination of diagnosis and/or causation involving conditions such as dementia, infectious disease, traumatic brain injury, cerebrovascular disease, toxic exposure, and general cognitive complaints. The mean age of this sample was 47.9 years with a standard deviation of 13.9 (range: 16 to 78) and 52% were female. Mean education was 13.6 years with a standard deviation of 2.8 (range: 6 to 20).

PROCEDURE

For each individual case, medical records were evaluated to determine if there was unequivocal evidence of structural brain abnormality. A case was considered to have established brain damage if the available neurodiagnostic (e.g., MRI, CT, EEG) findings were explicitly interpreted by a neuroradiologist or neurologist as being abnormal or, in a minority of cases, when radiology results were not reported but there was unequivocal evidence from clinical presentation of brain abnormality as determined by a neurologist (possibly based in part on neurodiagnostic data not available in the records). Table 1 describes the diagnostic composition of the brain damage sample. 18.1% of the sample ($N = 40$) was found to have brain damage by these criteria. Most of the cases classified as not having brain damage ($N = 180$) had cognitive complaints of unknown or multiple potential etiologies (80%), 12.8% had

Table 1 Diagnostic composition of the brain damage sample ($n = 40$)

	Evidence	
Diagnosis (n)	Radiologic	Clinical
Stroke	4	4
MTBI	3	0
M/S TBI	5	1
Dementia	3	5
Toxic	2	0
Multiple Sclerosis	3	3
Assorted neurologic	5	2
Total	25	15

MTBI = Mild traumatic brain injury; M/S TBI = moderate or severe traumatic brain injury.

a diagnosis of traumatic brain injury (TBI) of various severity, and 6.2% carried a diagnosis of some form of psychiatric disorder.

In addition, each case was classified as to whether external incentives were present for the evaluation. An external incentive was defined as being in a medical legal context (such as filing a disability or workers compensation claim or in a personal-injury lawsuit) or attempting to qualify for special benefits related to a neuropsychological condition. Of this sample 73.2% (47.5% of the brain damage sample, 78.9% of the non-brain-damage sample) had such external incentives and 26.8% did not. No criminal cases were included.

The results from two SVTs served as the major comparison variable. The Word Memory Test has been shown to be sensitive to poor effort and malingering in a wide variety of studies (Green, Allen, & Astner, 1996). It consists of the presentation of a list of words for the patient to remember. A forced choice method then presents one of the list's words and a foil, and the patient is then asked to identify which word was on the list. Accuracy of each response is given to the patient. For this study, only the "easy" subtests were utilized: Immediate Recognition, Delayed Recognition, and Consistency. Using the standards set forth in the manual and computerized program, cases were considered to have passed this test if the scores on each of these subtests exceeded 82.5%. All others were classified as having failed. The other SVT analyzed was the Computerized Test of Attention and Memory, a procedure that contains measures of both effort and genuine concentration problems developed by Fox (1990, 2009). The test uses a forced choice paradigm in which a 5-digit number (the stimulus) is very briefly shown on a computer screen. After a pause, two 5-digit numbers are presented and the test-taker indicates which of these two is identical to the stimulus. There are different levels of difficulty, and feedback is given by the computer regarding each response and general performance. Detectors of feigning have been derived from a wide variety of measures contained in the CTAM, including overall accuracy of performance, pattern of errors, and response time, as described in the manual. Using a variety of criteria and predictors, algorithms were developed based on multiple samples that yield a total Effort Score. Effort Scores less than 0 indicate poor effort, scores above 0 indicate adequate effort, and scores of exactly 0 are equivocal regarding the level of effort. Previous studies have demonstrated its effectiveness in detecting poor effort (e.g., Fox & Lees-Haley, 1995). Effort Scores below 0 have been found to have less than a 5% false positive rate in identifying cases that fail multiple other SVTs. The norms and scoring of the CTAM were not based on the current study sample. For this study, each profile was classified as having been passed if the Effort Score was greater than 0 and as having been failed if it was less than 0. A total of 27 cases showed equivocal effort (Effort Score = 0) on the CTAM and were dropped from subsequent analysis, yielding an *N* of 193.

The 24 neuropsychological tests analyzed are commonly used procedures for such evaluations although no standard battery was given in all cases. Age-corrected scores were computed for the Wechsler Adult Intelligence Scale III (WAIS) Full Scale IQ and the major indexes and the Wechsler Memory Scale III (WMS) indexes. Two scores thought to be sensitive to brain dysfunction were selected from the Stroop Color and Word Test (Golden & Freshwater, 2002): the Color-Word score (STROOP C-W) and the interference score (STROOP INT). The Wisconsin Card Sorting Test was administered and scored with the widely used commercially available WCST: CV4 computer program (Heaton & PAR Staff, 2005) and the categories completed (WCST CAT) and perseverative errors (WCST PE) scores were selected for analysis because they are reputed to be sensitive to brain damage (Lezak et al., 2004, p. 589). The Rey Complex Figure Test (RCFT) was administered and scored via procedures established by Meyers and Meyers (1995) and Kolb and Whishaw (1990). Scores consisted of the copy score (RCFT COPY), immediate memory (RCFT IMM), delayed memory (RCFT DEL), and percent retained (RCFT PERC). The remaining procedures (Controlled Oral Word Association (COWA), grooved pegboard for dominant (PEG DOM) and nondominant (PEG NDOM) hands, and Trail Making A (TRAILS A) and B (TRAILS B)) were scored via Heaton, Miller, Taylor, and Grant (2004). Scores for all tests were converted to standard scores (mean of 100, standard deviation of 15). Finally, because Green et al. (2001) found that SVT performance had a significant effect on the Overall Test Battery Mean (also described in Miller & Rohling, 2001), the mean standard score from all neuropsychological tests administered to an individual was calculated (Battery Mean).

RESULTS

Analysis shows the WMT and the CTAM results are correlated with each other ($r_{pb} = .66$, $p < .001$), as would be expected. The total sample failure rate was 34% for the WMT and 35% for the CTAM. Neither the WMT nor the CTAM performance was significantly correlated with brain damage ($r_{pb} = -.06$ for the WMT; $r_{pb} = .07$ for the CTAM). Interestingly, SVT performance was not correlated with incentive ($r_{pb} = -.04$ for the WMT; $r_{pb} = -.07$ for the CTAM).

Because some of the variables examined (e.g., brain damage present/absent and pass/fail of an SVT) are dichotomous and test performance is continuous, most correlation analyses utilized the point-biserial (Pearson's r_{pb}) statistic (DeCoster, 2004). The relationship between brain damage and the neuropsychological tests, differentiated by SVT performance, was investigated and the results are listed in Table 2. That table indicates that when the WMT is passed, 23 of the 25 correlations are statistically significant ($p < .05$, at least) in the predicted direction. Likewise, when the CTAM is passed, 22 of 25 test results are significantly correlated with brain damage in the expected direction. However, with failure on either SVT, none of the correlations reached significance. It is evident that, when an SVT is passed, there is the expected correlation between test performance and brain damage.

Table 2 Pearson point-biserial correlations (r_{pb}) between tests and brain damage for those passing or failing SVTs

	WMT				CTAM			
	Passed		Failed		Passed		Failed	
Test	*n*	r_{pb}	*n*	r_{pb}	*n*	r_{pb}	*n*	r_{pb}
COWA	130	−.16*	60	.04	108	−.15	59	−.04
PEG DOM	122	−.31**	58	.10	104	−.28**	56	−.06
PEG NDOM	121	−.31**	58	.11	103	−.29**	56	−.07
RCFT COPY	125	−.25**	69	−.06	106	−.29**	63	.00
RCFT IMM	125	−.30**	68	.06	106	−.30**	62	−.11
RCFT DEL	125	−.31**	69	.03	106	−.32**	63	−.11
RCFT PERC	125	−.30**	69	.05	106	−.27**	63	−.10
STROOP C–W	136	−0.19**	63	.09	114	−.17*	60	−.02
STROOP INT	137	−.09	63	.11	115	−.07	60	.04
TRAILS A	146	−.29**	74	.01	125	−.33**	68	−.09
TRAILS B	146	−.33**	73	.09	124	−.37**	68	−.03
WAIS FSIQ	130	−.29**	60	.06	110	−.33**	56	−.06
WAIS VCI	121	−.16*	54	.13	102	−.15	50	.00
WAIS POI	131	−.26**	59	.01	114	−.28**	52	−.13
WAIS WMI	136	−.33**	59	.21	116	−.38**	55	.11
WAIS PSI	137	−.26**	65	−.06	119	−.31**	57	−.18
WMS AUD IMM	133	−.17*	59	−.09	111	−.20*	55	−.19
WMS VIS IMM	125	−.22**	59	.01	108	−.28**	52	−.11
WMS IMM	118	−.19*	50	−.08	97	−.28**	48	−.20
WMS AUD DEL	133	−.16*	59	−.11	111	−.20*	55	−.19
WMS VIS DEL	124	−.24**	59	.08	107	−.29**	52	−.11
WMS GEN	116	−.22**	49	−.08	95	−.24**	47	−.22
WCST PE	110	−.13	40	−.10	94	−.24**	38	.10
WCST CAT	107	−.21**	40	−.07	92	−.32*	37	−.03
Battery Mean	146	−.40**	74	.03	125	−.44**	68	−.09

COWA = Controlled Oral Word Association; PEG = Grooved Pegboard: DOM = dominant hand, NDOM = non-dominant hand; RCFT = Rey Complex Figure Test: COPY = copy, IMM = immediate recall, DEL = delayed recall; PERC = percentage recalled; STROOP = Stroop Color and Word Test: C-W = color-word, INT = interference; TRAILS A = Trail Making, part A; TRAILS B = Trail Making, part B; WAIS = Wechsler Adult Intelligence Scale III: FSIQ = Full Scale IQ, VCI = Verbal Comprehension Index, POI = Perceptual Organization Index; WMI = Working Memory Index: PSI = Processing Speed Index, WMS = Wechsler Memory Scale III, AUD IMM = Auditory Immediate Index, VIS IMM = Visual Immediate Index, IMM = Immediate Index, AUD DEL = Auditory Delayed Index, VIS DEL = Visual Delayed Index, GEN = General Memory Index; WCST = Wisconsin Card Sorting Test: PE = perseverative errors, CAT = categories completed. See text.

*$p < .05$, one-tailed. **$p < .01$, one-tailed. All others *ns*.

Table 3 Average battery mean and effect sizes for SVT performance and brain damage

Variable	*N*	Mean (*SD*)	Cohen's *d*	*r*
WMT	220		1.22	.52
Pass	146	93.65 (12.02)		
Fail	74	78.18 (14.16)		
CTAM	193		1.01	.45
Pass	125	93.79 (12.05)		
Fail	68	79.89 (14.63)		
Brain Damage	220		0.64	.30
Present	40	81.34 (13.25)		
Absent	180	90.83 (14.58)		
Brain Damage[a]	112		1.30	.54
Present	22	82.48 (12.84)		
Absent	90	97.71 (10.50)		
Total	220	88.90 (15.04)		

[a]Cases in which both the CTAM and WMT were passed.

Failure on an SVT eliminates the relationship between neuropsychological test performance and brain damage. This effect is particularly pronounced for the test Battery Mean, which shows the highest correlation with brain damage among the tests when an SVT is passed but absolutely no relationship when it is failed. The relative influence of SVT performance and brain damage on the Battery Mean is further illustrated in Table 3 that describes effect sizes using Cohen's *d*. There is a very large effect size for the WMT, a large effect for CTAM and a medium effect for brain damage.

Additional analysis was conducted using multiple regression procedures without intercept (to elucidate the contribution of each factor) to predict the Battery Mean using the WMT, CTAM, and brain damage variables as predictors. Considered separately, the WMT accounted for a large portion of the variance ($R = .85$, $R^2 = .72$), with the CTAM performing comparably ($R = .84$, $R^2 = .70$). Brain damage produced an R of .38 (R^2 of .15). When the variables were entered in a stepwise fashion, the total obtained R^2 was .77 ($n = 193$). The SVT variables reached significance ($p < .001$) with 73% of the variance accounted for by the WMT and an additional 3.1% by the CTAM. Brain damage also reached significance ($p < .05$) and accounted for an additional .6% of the variance. Of note, when cases were selected in which both the CTAM and WMT were passed ($n = 112$), the effect size for brain damage was very large (see Table 3).

DISCUSSION

These results confirm all four major hypotheses. In particular, failure on an SVT essentially invalidates the relationship between neuropsychological test results and brain damage. When even one SVT is failed, there is no correlation between cognitive test performance and documented brain damage. These data thus successfully respond to the challenges raised by McGrath et al. (2010) and demonstrate that SVTs can indeed accurately measure a response bias that directly influences neuropsychological test validity. In this case the data validate both the WMT and CTAM as indicators of response bias in neuropsychological assessment. Likewise, effort as measured by SVT performance has a greater effect on test scores than does brain damage in this general neurological population, some of whom did not have external incentives. In fact, effort as measured by either the WMT or the CTAM, accounts for nearly five times as much variance as does brain damage. These findings are consistent with the results of Green et al. (2001). The data also demonstrate that once symptom validity is established, brain damage has a very large effect on neuropsychological test results. Neither the WMT nor the CTAM was affected by brain damage, which means that failure on such tests cannot be explained by the presence of brain damage, at least in the neuropsychological population represented in this sample. About 35% of these examinees failed the WMT or CTAM, a rate that is comparable to previous estimates of the rate of malingering and poor effort (Larrabee, 2003). Surprisingly, SVT performance was not associated with incentive in this sample. These findings thus lend further support to the increasing practice of giving SVT tests to all examinees regardless of whether incentive is present. They reinforce the necessity of using SVT methodology and the conclusion that failure on even one accepted SVT can invalidate the findings from neuropsychological batteries.

There are some limitations to these findings. It is acknowledged that the criteria for brain damage in this study are imprecise. However, it is unbiased with regard to passing or failing an SVT, which served as the major grouping variable. Different or more detailed criteria for brain damage may yield different correlations but the discrepancies in these correlations between passing and failing an SVT are likely to remain. For an SVT to be considered effective it should measure the validity of neuropsychological tests, something that has been demonstrated here for the WMT and CTAM. This technique of illustrating the effect of SVT failure is a promising procedure for establishing the construct validity of proposed SVTs as well. Future studies using this paradigm should be conducted with additional samples of SVTs, neuropsychological tests and neuropathology.

DISCLOSURE

The author is also the author of one of the tests (CTAM) utilized in this study.

ACKNOWLEDGMENT

My appreciation goes to Paul Lees-Haley Ph.D. for his helpful comments on this research.

REFERENCES

Bigler, E. D. (2001). The lesion(s) in traumatic brain injury: Implications for clinical neuropsychology. *Archives of Clinical Neuropsychology*, *16*, 95–131.

Boone, K. (Ed.). (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: The Guilford Press.

Bush, S., Ruff, M., Troster, I., Barth, J., Koffler, S., Pliskin, N., et al. (2005). NAN position paper. Symptom validity assessment: Practice issues and medical necessity. Nan Policy and Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 419–426.

DeCoster, J. (2004). *Data analysis in SPSS*. Retrieved August 9, 2010 from http://www.stat-help.com/notes.html

Fox, D. D. (1990, 2009). *Manual for the Computerized Test of Attention and Memory* (8th ed.). Glendale, CA: D. D. Fox.

Fox, D., & Lees-Haley, P. (1995, June). *Preliminary validation of a new test of malingered memory problems: The CTAM*. Presented at the Computerized Behavioral Testing of Humans in Neurotoxicology Research. Portland, OR.

Golden, C., & Freshwater, S. (2002). *The Stroop Color and Word Test Manual*. Wood Dale, IL: Stoelting Co.

Green, P. (2007). Spoiled for choice: Making comparisons between forced choice effort tests. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 50–77). New York: The Guilford Press.

Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test* [manual]: *A user's guide to the oral and computer-administered forms* (US Version 1.1). Durham, NC: CogniSyst, Inc.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery:Demographically adjusted neuropsychological norms for African-American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources Inc.

Heaton, R., & Staff., P.A. R. (2005). *WCST: Computer version 4 – Research edition [computer software]*. Lutz, FL: Psychological Assessment Resources.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*, 1093–1129.

Kolb, B., & Whishaw, I. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: W. H. Freeman & Co.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–425.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment* (4th ed.). Oxford: New York.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, *136*(3), 450–470.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and recognition trial: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Miller, L. S., & Rohling, M. L. (2001). A statistical interpretive method for neuropsychological test data. *Neuropsychology Review*, *11*, 143–169.