OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

Archives of Clinical Neuropsychology 26 (2011) 8–15

Archives of CLINICAL NEUROPSYCHOLOGY

Embedded Symptom Validity Tests and Overall Neuropsychological Test Performance

John E. Meyers[1,*], Marie Volbrecht[2], Bradley N. Axelrod[3], Lorrie Reinsch-Boothby[1]

[1] *Meyers Neuropsychological Services, Mililani, HI, USA*
[2] *Neuropsychology Consultants, Sioux Falls, SD, USA*
[3] *John D. Dingell Department of Veterans Affairs Medical Center, Detroit, MI, USA*

*Corresponding author at: John E. Meyers, Meyers Neuropsychological Services, Mililani, HI, USA. Tel.: 712-251-7545.
E-mail address: jmeyersneuro@yahoo.com (J.E. Meyers).

Accepted 17 October 2010

Abstract

A sample of 314 consecutive clinical and forensic referrals with mild traumatic brain injury was evaluated using the Meyers Neuropsychological Battery (MNB). A comparison was made of the test performance and performance on the embedded Symptom Validity Tests (SVTs) with a control for multicolinearity utilized. Using the nine embedded SVTs in the MNB, the incidence of poor effort fell at 26% of the total sample. Involvement in litigation was related to more failures on the individual SVTs. The correlation between failed effort measures and the Overall Test Battery Mean (OTBM) was consistently negative, regardless of litigation status, in that more failures were associated with lower OTBM scores. The correlation between the number of SVTs failed and the OTBM was −.77. Our results are similar to those presented by Green, Rohling, Lees-Haley, and Allen (2001); who reported a .73 correlation with the failure on the Word Memory Test and performance on the OTBM. The results of the current study also indicate that 50% of the variance in neuropsychological testing can be accounted by failures on internal SVTs.

Keywords: Malingering/symptom validity testing; Meyers Neuropsychological Battery; Overall Test Battery Mean

Internal and add-on Symptom Validity Tests (SVTs) are quite commonly used in neuropsychological assessments to objectively evaluate diminished effort (Larrabee, 2003, 2008; Meyers & Volbrecht, 2003). In fact, the American Academy of Clinical Neuropsychology recommends explicit evaluation of symptom validity (Heilbronner et al., 2009). In addition, The National Academy of Neuropsychology presented a position paper on the use of SVTs (Bush et al., 2005), which specifically recommends administration of at least two SVTs in an evaluation.

In addition to the external SVT tests such as the Word Memory Test (WMT), which are added onto a battery of tests, neuropsychologists often use internal SVTs that are built into the tests given as part of a battery. The use of these internal or embedded SVTs is now a common approach in neuropsychology (e.g., Bush et al., 2005). An internal SVT is a measure assessing the validity of test performance based on the tests already given as part of the battery of tests. These internal SVTs can be based on a variety of methods for assessing validity. For example improbably poor performance is based on performance that is lower than a predetermined cutoff score. Another method employed is the inspection of relationships or consistency between scores. Meyers (2007) found that the approach used by those asked to simulate malingering on neuropsychological tests varied based on the nature of the individual test presented. In addition, different approaches may be used by different individuals when attempting to "malinger." The use of internal SVTs allows the neuropsycologist to assess for a wider variety of approaches that one might use in choosing to perform poorly on neuropsychological test (Meyers 2007).

The Meyers Neuropsychological Battery (MNB; Meyers & Rohling, 2004) includes not only measures of cognition, but also tasks of symptom validity and is a well-researched and established battery of tests. The MNB is composed of commonly used neuropsychological measures; the MNB assesses domains of attention and working memory, processing speed and mental

doi:10.1093/arclin/acq083

flexibility, verbal reasoning, visual reasoning, verbal memory, visual memory, dominant motor and sensory, and nondominant motor and sensory. These domains of cognition are common neuropsychological constructs (Strauss, Sherman & Spreen, 2006).

The MNB has been used in previous studies. Rohling, Meyers, and Millis (2003) found that the length of loss of consciousness (LOC) was related to the level of expected cognitive impairment as measured by the mean of the test scores on the battery of test; this is referred to as the Overall Test Battery Mean (OTBM) of the MNB and is expressed in a *t*-score metric. The results of that 2003 study were nearly identical to those presented by Dikmen, Machamer, Winn, and Temkin (1995), using an expanded Halstead Reitan Battery. The results of these two studies indicate that there is a consistent relationship between the level of OTBM and the length of LOC. Rohling and colleagues (2003) found that for those with a length of LOC less than an hour, the average OTBM was 44.3 ($SD = 5.0$), with progressively lower OTBM found as the length of LOC increased. The findings of this study and that of Dikmen and colleagues (1995) indicate that there is a relationship between LOC and the expected performance level as measured by the OTBM. The results of these two studies indicate that mild TBI injuries are expected to show only mild deficits on neuropsychological testing, and that LOC has a relationship to overall neuropsychological test performance. Therefore, individuals with similar LOC would be expected to score at, generally, the same level of impairment. Given these consistent findings of a relationship between LOC and OTBM, the LOC was used as a variable for the current study to identify the level of severity of injury.

In a comprehensive neuropsychological test battery, Green, Rohling, Lees-Haley, and Allen (2001) found that efforts accounted for more of the variance than did injury severity. Individuals were given the WMT (Green, 2003), an independent add-on symptom validity measure. They reported a significant correlation between the WMT and OTBM (of their flexible neuropsychological battery) of .73. Individuals who failed the WMT performed worse on neuropsychological tests compared with those who passed the WMT.

The MNB uses built-in SVTs which have been validated in previous publications (Meyers, Morrison & Miller, 2001; Meyers & Volbrecht, 1998, 1999; Meyers, Galinsky & Volbrecht, 1999; Meyers, Bayles & Meyers, 1996), the cutoffs for each individual SVT contained in the MNB were established, and then the second validation of these cutoffs was presented in Meyers and Volbrecht (2003). The cutoff scores for each of the MNB SVTs were individually set at a zero false positive (FP) rate. Then Meyers and Volbrecht (2003) found that, when the nine individual SVTs in the MNB were used in combination, they were able to differentiate valid from invalid test performance. In other words, previous research has identified the level of performance at which no one is erroneously identified as malingering. Then using the SVTs as a group, Meyers and Volbrecht found that a failure rate of two or more of the individual SVTs (that were set at a zero FP rate) were able to identify a data set as invalid. Obtaining only one failure on the nine validity checks is not indicative of deficient effort. The criterion of using two individual SVT failures to indicate invalid test performance was validated in a study by Meyers and Volbrecht (2003). Larrabee (2008) recently found that a cutoff score of three SVT failures (when using SVTs with more than a zero FP rate) would suggest invalid test performance (Larrabee, 2008). Using a large clinical sample of 796 participants, it was found that the nine internal validity checks when used together were able to identify litigants from non-litigant groups consisting of mild, moderate, and severe traumatic brain injured (TBI) patients as well as chronic pain, depressed, and community controls. Using the criteria of two or more internal SVT failures to define poor effort resulted in an 83% sensitivity and a 100% specificity overall. Internal SVTs often use different methods of identifying “invalid” test performance (Meyers, 2007). Some internal SVTs use improbable low performance, while others use relationships between scores as measures of consistency (Meyers, 2007).

Some individuals may argue that the major weakness of using built-in SVTs is that the tests themselves have been found to be sensitive to brain injury and therefore the tests can be failed due to cognitive difficulties related to brain injury. However, the study by Meyers and Volbrecht (2003) clearly found that, if an individual was not living in an institutional setting, there should be zero–1 SVT failures. When internal SVTs are used appropriately with the populations they are designed for, the SVT may be used as both a cognitive measure and as an SVT. The advantage of internal SVTs is the ability to assess the validity of performance on the actual tests used to assess cognition. In addition, internal SVTs do not add additional time to the testing time; this is an important benefit in this time of managed care. Although not all tests have built-in SVTs, this is a desirable goal.

As was previously discussed, Green (2003) found a significant correlation between the WMT and the OTBM of .73. A review of literature reveals no similar studies using internal SVTs methods. The purpose of this study was to identify whether poor performance on embedded SVTs also results in poor performance on other neuropsychological tests. This study is not a validation study of the internal SVT, which has already been established in the studies already cited in this article. This study examines the relationship of poor performance on internal SVTs and the effect of this poor performance on a battery of neuropsychological tests. It is hypothesized that there will be a relationship between performance on internal SVTs and overall performance on a battery of neuropsychological tests as identified by the OTBM.

Method

Participants

Three hundred fourteen consecutive outpatient referrals (170 were referrals by attorneys/case managers, the remainder were referrals by physicians), all of whom reported a head injury with lasting symptoms, were considered for inclusion. The criteria for final inclusion in this study were:

(1) Individuals with a reported LOC for 5 min. This included persons who reported being briefly "dazed," "confused," or had a brief loss of awareness or similar description at the time of the injury. All were claiming mild cognitive difficulties.
(2) All individuals included in this study had a normal magnetic resonance imaging (MRI) prior to the assessment.
(3) All were uncomplicated injury events with no additional orthopedic or other injuries.
(4) All individuals where community-dwelling individuals who were independent in all daily activities, including driving.
(5) Any individual with missing test data was excluded from the study.

For the individuals included in the study, Glasgow Coma Scale or length of post traumatic amnesia were not consistently available and so could not be used in this study. Information on any single proton emission computed tomography or proton emission tomography scans was also not consistently available and thus not used as a defining variable for this study. All individuals had MRI information available in their medical records; however, detailed descriptions of the protocols used for each of the individuals were not consistently available and so were not available for each participant. Individuals included in this study had at least one MRI prior to the neuropsychological assessment with some individuals having more than one MRIs (including what may have been done at the time of injury). For those that did have an MRI or computerized tomography scan at the time of injury, all were read by a neuroradiologist as indicating no intracranial abnormalities. The length of LOC was based on the participants' own report as recorded at the time of the ambulance or Emergency Room admission in his/her medical records; due to the inherent difficulty of not having a method of accurately identifying the length of LOC, the variable is simply coded present or not in order for the data to be used in this study, the length of reported LOC had to be <5 min. In all cases that LOC was referred to at the injury scene, records indicated that the individual was awake and alert at the time the ambulance or police arrived.

It is expected that individuals with mild brain injuries would pass these SVTs, as the criteria for failing these SVTs were established on individuals who have severe TBI (Meyers & Volbrecht 2003). In this study, we chose to examine the individuals with mild TBI; in this way we could control the effects of different levels of injury by restricting the sample to only mildly injured individuals, based on their LOC. This mild TBI population is also the patient group that usually has the most questions regarding SVT performance, since Mild TBI groups tend to fail SVTs more often compared their moderate to severe TBI counterparts (Green 2003), they represent an obvious group of interest to study. Therefore, a mild TBI group was selected using the criteria listed above.

Measures

The individual tests that make up the MNB (in the order of administration) are:

WAIS-III (Wechsler, 1997), Picture Completion, Digit Symbol, Similarities, Block Design, Arithmetic, Digit Span, Information, Forced Choice (FC), Rey Complex Figure and Recognition Trial (CFT) Copy (Meyers & Meyers 1995), Animal Naming (Strauss et al., 2006), 1 min Estimation (MNB electronic Manual), CFT Immediate delay (Meyers & Meyers, 1995), Controlled Oral Word Association Test (Strauss et al., 2006), Dichotic Listening (Meyers, Roberts, Bayless, Volkert & Evitts, 2002), Sentence Repetition (SR; Strauss et al., 2006), CFT 30 min delay (Meyers & Meyers, 1995), CFT Recognition (Meyers & Meyers, 1995), Rey Auditory Verbal Learning Test (AVLT) Trial 1 (Strauss et al., 2006), AVLT Total (Trials 1–5) (Strauss et al., 2006), AVLT Immediate delay Strauss et al., 2006), Judgment of Line (JOL) Orientation (JOLO; Benton, Hamsher, Varney & Spreen, 1983). Boston Naming (Strauss et al., 2006), Finger Tapping (Reitan & Wolfson, 1985), Finger Localization (Benton et al., 1983), Trails A (Reitan, & Wolfson, 1985), Trails B (Reitan & Wolfson, 1985), Token Test (Strauss et al., 2006), AVLT 30 min delay (Strauss et al., 2006), AVLT recognition Strauss et al., 2006), Category Test—Victoria Version (Kozel & Meyers,1998).

The internal SVTs are:

(1) Reliable Digits (RLDS) is calculated by taking the longest span of Digit Span forward on which both trials are passed plus the longest span on Digit Span backwards on which both trials are passed. Several studies have examined the use of RLDS as an SVT. Meyers and Volbrecht (1998, 2003) proposed a cutoff of 6 or below for a zero FP rate. Greiffenstein, Baker, and Gola (1994) proposed a cutoff of 7, whereas a study by Greve and colleagues (2009) found that a cutoff of 6 or less separated good versus invalid test performance. For the MNB validity measure, a score of 6 or below is a considered a failure on the RLDS. For purposes of this study, we used the WAIS-III version of Digit Span.
(2) JOL is the raw uncorrected number correct on JOL with a score of 12 or below considered a failure on the SVT (Meyers et al., 1999; Meyers & Volbrecht, 2003).
(3) Dichotic Listening-Both (DLB) is the score on the dichotic listening task for both ears. A score of 9 or less indicates a failure on this SVT (Meyers & Volbrecht, 2003). DLB is not used if the examinee does not have sufficient hearing for the task. A hearing test is administered at the beginning of the task.
(4) Token Test (TT) (Strauss et al., 2006) is scored by the number of points correct on the TT task. A score of 150 or less indicates a failure on this SVT (Meyers & Volbrecht 2003).
(5) SR (Strauss et al., 2006) is scored as the number of sentences that were repeated correctly. A score of 9 or below indicates a failure on the SVT (Meyers & Volbrecht, 2003).
(6) AVLT –Recognition (AVLT-Rec) (Strauss et al., 2006) is scored as the number of True Positive responses on the AVLT Recognition. A score of 9 or below is a failure on this SVT (Meyers & Volbrecht, 2003).
(7) Memory Error Pattern (MEP) uses the relationship of the immediate, delayed, and recognition scores on the CFT. The profiles for this SVT are based on these relationships and are delineated Attention, Encoding, Storage, Retrieval, Consolidation, Peak, and Other (Meyers & Volbrecht, 1999, 2003). Production of an Attention, Encoding or Storage MEP indicates failure on the SVT. These MEPs of Attention, Encoding and Storage are only expected with individuals who are institutionalized due to cognitive impairment. An Attention MEP is classically identified as scores on the CFT immediate, delayed, and recognition scores all below a score of 20 on a *t*-score scale (Meyers & Meyers, 1995). An encoding MEP is when the CFT immediate and delayed are $<$20 on a *t*-score scale and recognition is higher, (usually $>$20 on a *t*-score scale) (Meyers & Meyers, 1995). The storage MEP is a downward slope when CFT immediate $>$ CFT delayed $>$ recognition.
(8) FC (Brandt, Rubinsky, & Larson, 1985) is a 20-item FC recognition task. A score of 10 or below is considered as a failure on the SVT (Meyers et al., 1999; Meyers & Volbrecht, 2003).
(9) Estimated Finger Tapping is calculated using tests that employ the dominant hand fingers to perform the test being compared with the mean raw score of the Finger Tapping test (Meyers & Volbrecht, 2003). The formula used is estimated FT = ((Block Design scale score) × 0.361) + (Digit Symbol scale score) × 0.491) + (raw Copy score on CFT) × 0.185) + 31.34). Difference = estimated FT – actual mean finger tapping score for the dominant hand. If the difference is $>$10, then this is a failure on the SVT.

In conclusion, the criteria for pathological performance on the MNB SVTs include: RLDS ($\leq$6), JOLO ($\leq$12), Dichotic Listening-Both (DL-B $\leq$9), TT ($\leq$150), SR ($\leq$9), AVLT recognition (AVLT-Rec $\leq$9), MEP for the CFT (MEP: Attention, Storage or Encoding); FC ($\leq$10) and relationship between actual performance on finger tapping and the expected finger tapping performance based on performance on other tests. Failures on two or more of these SVTs are considered a failure on the Meyers Validity Index (Meyers & Volbrecht, 2003) and indicate an invalid data set.

Statistical Analysis

All tests scores on the individual tests on the MNB were converted to *t*-scores, using the individual norms for each test. To avoid the difficulty of multicolinearity, those measures in the MNB that contain the SVTs of interest were separated out. The OTBM was calculated as the mean of the *t*-scores for the 19 tests that were left after the SVT tests were removed.

Due to the known relationship between SVT performance and litigation status (Green et al., 2001), the first analysis was a *t*-test analysis of litigation status and an examination of demographic variables, including LOC. Means and standard deviations of the litigants and non-litigants were calculated, as well as the mean performance on the OTBM based on the number of SVT failures. An independent samples *t*-test will be computed to compare the litigants and non-litigants on SVT failures and OTBM performance. The data were then recombined and were then separated by the number of SVT failures. Means, standard deviations, and effect sizes were calculated along with *t*-test comparing the number of SVT failures and OTBM performance. Statistical comparisons will also be presented on demographic variables, SVT, and OTBM with litigation status.

Results

Of the sample of 314 individuals, the mean age of the sample was 39.2 ($SD = 12.9$) and years of education was 12.4 ($SD =$ 2.4). Other sample characteristics are that 276 were right handed and 38 were left handed; 302 were Caucasian, 1 identified himself other, 2 were African American, 3 were Native American, 6 were Hispanic; 111 were female and 203 were male. At the time of the assessment, 170 were in litigation or disability proceedings, while 144 were not in litigation.

Those in litigation were compared with those not involved in litigation to investigate any impact this status may have had on results. The litigants were older and less educated than the non-litigants. The mean age was 41.6 ($SD = 11.6$) years for litigants and the mean age was 36.4 ($SD = 13.7$) for nonlitigants. The mean years of education was 12.2 years ($SD = 2.3$) for the litigants and 12.7 ($SD = 2.5$) for the nonlitigants. There were no significant differences, using independent samples *t*-test, for both age and education (p's $> .05$). The groups were then separated based on litigation status. In order to ascertain if this difference was meaningful to the study at hand, a Pearson correlation was computed, comparing the age and education with the number of SVT failures. Given the non-normal distribution of the SVT data, a Fisher's r' transformation was completed. The small value of the correlations resulted in the transformation being equivalent to the values of the Pearson *r*, suggesting that the Pearson value is sufficient for this data. The results indicated that the correlation was .11 ($P > .05$) with age and SVT failures and $-.13$ ($p > .05$) with education. These scores indicated that, for litigants, age and education were not significantly related to SVT performance. The same analysis was then performed for the nonlitigant group. The correlations when age and education were compared with the number of SVTs failed were .12 and $-.03$, respectively, and neither was significant (p's $> .05$). Table 1 shows the mean number of SVT failures and OTBMs for both the litigants and nonlitigants.

Both the litigant and the nonlitigant groups were combined for all the remainder of the analyses. The correlation between litigation status, age and education was calculated using a Pearson *r*. As litigation status is dichotomous, a point biserial correlation was originally considered. However, a "Pearson's *r* in this case is mathematically equivalent to a point biserial *r*" (Andrews, Klem, Davidson, O'Malley, & Rodgers, 1981, p. 7). Litigation status correlated with age at $-.20$ and with education at .11. The correlation was negative for age, indicating that as age increased, the number of SVT failures decreased. Therefore, although these differences were statistically significant, they account for very little variance with age accounting for 4.1% and education accounting for only 1.2%. Therefore, as the amount of variance accounted for is quite small, it is likely to be of little to no clinical importance. A chi-square analysis indicated that litigation status and gender are not strongly related, χ^2 $(1, N = 314) = .04$, $p = .90$, neither is handedness , $\chi^2(1, N = 314) = 1.20$, $p = .54$, and neither is ethnicity χ^2 $(4, N = 314) = 3.577$, $p = .46$.

Using the same cutoffs as reported by Meyers and Volbrecht| (2003) of the sample of 314 individuals, 232 had either no or only one failure on the SVTs. The remaining 26% of the sample presented with two or more failed SVTs, which is considered indicative of poor effort on the test battery. OTBM performance relative to the number of SVTs failed on the MNB is presented in Table 2. As is evident, lower OTBM scores were often generated by individuals who produced more failures on the SVT measures.

The LOC variable was coded a zero for those who reported only being dazed (with no LOC) and a 0.01 for those who reported a brief LOC with some loss of time awareness $<$5 min. As most neuropsychologists would agree, the exact measurement of the LOC is not possible without an external monitoring of some short. None of the individuals in this study had an outside monitoring of their confusion/LOC. Given this limitation of the data available, only categorical data were available for use. Results showed that the LOC did have a relationship to the number of SVTs failed (χ^2 $(8, N = 314) = 35.246$, $p < .001$). Of those who reported no LOC, 118 of 227 (54%) failed one or more SVTs; whereas of those who reported having some LOC, only 31 of 97 (31%) failed one or more SVTs. Further investigation revealed that the mean OTBM performance for those who reported an LOC was 45.5 ($SD = 7.0$), and for those who reported no LOC was 41.2 ($SD = 9.1$). These results indicate that those who reported an LOC actually scored higher on the OTBM than did those who did not report LOC. Given these results, one could observe that the absence of LOC has a relationship with SVT performance. That is, no LOC results in more SVT failures. Therefore, clinically it is clear that the presence of LOC does not, in and of itself, produce more SVT failures.

Table 1. Mean (and standard deviations) of MNB SVT failures and OTBM by litigation status

	Litigation Pending ($n = 170$)	No Litigation Pending ($n = 144$)	*Cohen's *d*	*t*-test
MNB SVT failures	1.6 (1.9)	.45 (0.90)	.77	$t = 6.810$, $p < .001$
OTBM on MNB	39.8 (8.9)	45.8 (7.3)	.73	$t = -6.433$, $p < .001$

Note: SVT = Symptom Validity Tests; MNB = Meyers Neuropsychological Battery; OTBM = Overall Test Battery Mean.
*Guidelines used in interpreting effect size: negligible $\leq$.19; small = .20 to .49; medium = .50 to .79; large $\geq$.80 (Cohen, 1988).

Table 2. Mean (and standard deviations) of OTBM performance by the number of MNB SVTs failed

No. of SVT Failures	OTBM (SD)	Comparing no. of SVT failures	Cohen's d	t-test, significance
0 ($n = 165$)	48.2 (4.4)	0 and 1	1.32	$T = 9.760$, $p < .001$
1 ($n = 67$)	41.3 (5.9)	1 and 2	0.76	$T = 3.485$, $p = .001$
2 ($n = 30$)	36.7 (6.1)	2 and 3	0.97	$t = 3.297$, $p = .002$
3 ($n = 20$)	31.2 (5.1)	3 and 4	−0.01	$t = -.041$, $p = .968$
4 ($n = 13$)	31.3 (8.9)	4 and 5	0.05	$t = .099$, $p = .922$
5 ($n = 10$)	30.9 (4.6)	*		
6 ($n = 5$)	22.9 (4.7)	*		
7 ($n = 2$)	20.8 (5.5)	*		
8 ($n = 2$)	25.5 (5.9)	*		
9 ($n = 0$)				

Notes: SVT = Symptom Validity Tests; MNB = Meyers Neuropsychological Battery; OTBM = Overall Test Battery Mean; **n* too small for reliable calculations; Guidelines used in interpreting effect size: negligible ≤.19; small = .20 to .49; medium = .50 to .79; large ≥.80 (Cohen, 1988).

Table 3. Pearson correlations of cognitive domain with the number of MNB SVTs failed (all correlations were significant $p < .001$)

Cognitive domain	R
OTBM	−.77
Attention and working memory	−.45
Processing speed and mental flexibility	−.75
Verbal reasoning	−.77
Visual reasoning	−.68
Verbal memory	−.65
Visual memory	−.65
Dominant motor and sensory	−.65
Nondominant motor and sensory	−.48

Note: These correlations are all inverse correlations of MNB SVT Failures (0 to 9) and each cognitive domain.
SVT = Symptom Validity Tests; MNB = Meyers Neuropsychological Battery; OTBM = Overall Test Battery Mean.

In looking at the data in Table 2, it is observed that the OTBM performance on the MNB declined as the number of SVTs failed increased. Cohen's *d* was calculated for comparisons of the number of individual SVTs failed. One can see that there is a notable and strong effect of SVT failures on test performance with a failure of one SVT, with another large effect with failure of two SVTs. In fact, OTBM scores were significantly higher for individuals who failed none or only one SVT ($M = 45.8$, $SD = 6.0$) in comparison with those who failed two or more SVTs ($M = 30.8$, $SD = 6.7$); the difference between these two scores is significant and with a Cohen's $d = 2.35$. The OTBM score for all groups who failed two or more SVTs fell <1.5 standard deviations below the mean test performance. When three SVTs are failed, there is also a strong effect; however, after three failures, the effect is less pronounced. Comparisons with six or more failures could not be reliably calculated due to the small number of individuals generating so many failures. Clearly there is a strong effect on test performance with SVT failures.

The correlation between the number of SVTs failed and the OTBM was significant (see Table 3), accounting for 59% of the variance. The results of examination of the specific domains relative to SVT failures also are given in Table 3. Strong correlations were also observed on tasks of processing speed and reasoning (both verbal and nonverbal); all correlations were significant ($p < .001$); suggesting a relationship between the number of SVT failures and the OTBM.

The data in Table 4 show the frequency of SVT failures. It is observed that AVLT—recognition was the most common SVT to be failed (20%). This is a verbal memory-based SVT. Memory-based SVT is a common approach to "malingering" (Green et al., 2001; Meyers, 2007); as such, a higher rate of SVT failures on this type of SVT is to be expected.

Discussion

Results indicate that performances on the embedded SVTs included in the MNB are related to OTBM performance. Specifically, the more SVTs failed, the lower the OTBM (even when the internal SVTs were removed from the OTBM calculation). The results of this study show that the correlation between internal SVT failures and the OTBM was 0.77; Green and colleagues (2001) reported a correlation of 0.73; the results of these two studies are similar. Another finding to this

Table 4. Frequency of individual SVT failure

Internal SVTs	No. of failures
AVLT recognition	64 (20%)
Reliable digits	56 (17%)
Dichotic listening both	52 (16%)
Estimated finger tapping	51 (16%)
Memory error pattern	41 (13%)
JOL orientation	34 (10%)
Sentence repetition	27 (8%)
Forced choice	21 (6%)
Token test	20 (6%)

study is that, in general, litigants do seem to fail SVTs more than nonlitigants, and thus have lower OTBMs. This again is consistent with the findings of Green and colleagues (2001).

An examination of SVT failures showed a relationship to LOC. Our results showed that those with a reported LOC actually performed better overall on the OTBM, and had fewer SVT failures compared with those who did not report an LOC. We were able to conclude that cognitive impairment associated with reported LOC was not the cause of the SVT failures in the sample used in this study. As was previously discussed, the SVT failures were not strongly related to other demographic variables. There as a tendency for older, less-educated individuals to fail SVT, but this is likely an incidental finding to our sample. It may be that older, less-educated persons do malinger more often, as they may feel they have fewer options and so may malinger more. It is possible that the perception of fewer choices may motivate one to "malinger" more often. This would be the topic for further investigation. The authors conclude that factors related to motivation (and not the SVT items themselves) are likely associated with these SVT failures, which is consistent with the conclusion presented by Green and colleagues (2001).

The recommendations of Bush and colleagues (2005) indicate that motivation and effort should be explicitly assessed as part of all neuropsychological assessments. One cannot assume that an individual in litigation is presenting with poor motivation/effort or that one not involved in litigation is offering a valid presentation. One must explicitly assess for effort using well-established SVTs. The results of the current study also support this recommendation. The importance of the finding of this study is that there does appear to be a relationship between internal SVT failures and overall neuropsychological test performance. Poor performance on SVTs correlates with poor performance on neuropsychological tests. The current study uses the built-in SVTs included in the MNB, and it is likely that other neuropsychological tests not included in this battery of tests will also show a relationship with SVT performance and further study is recommended.

In this current study, it would have been helpful to have more individuals who failed six or more of the SVTs. The current results indicate that no one failed all nine SVTs. These results indicate that SVT failure is not an all or none phenomenon. That is, individuals do not fail all SVTs equally, but instead are more selective of their SVT performance. These findings are consistent with those reported by Meyers (2007) who indicated that the approach to malingering varies based on the nature of the task and the "perception" of the individual as to what cognitive deficits he/she is trying to portray.

The number of SVT failures as identified in Table 2 show that there is a significant drop in OTBM between zero and one SVT failures, and also a significant drop between 1 and 2 SVT failures. Between two and three, there is also a significant drop in OTBM, but after two failures (i.e. three or more) there is no significant difference in the amount of decline in OTBM. These findings are similar to the previous findings of Meyers and Volbrecht (2003) and Larrabee (2008), and support that failure on two or more SVTs likely indicates an invalid data set.

This study used a specific set of tests that have been found to be capable of discriminating normal controls from individuals with mild traumatic brain injury. The results from this study may not generalize to all "flexible" batteries. Similarly, the SVTs used in this study have been specifically validated, and the cutoffs have specifically been set at a zero FP rate. The combination of the internal SVTs was set at a zero FP rate. When using other SVTs that do not have a zero FP rate, a different result may be obtained. Future studies addressing these findings with other batteries of tests and other internal SVTs are recommended. Finally, the dataset used in this study was comprised on individuals who had relatively brief LOC or confusion at the time if injury. It is possible that if there was a greater LOC, that the results may have been different. Therefore, future study could be directed toward the examination of internal SVT with individuals who have longer LOC.

The conclusions of this study indicate that OTBM performance on a battery of neuropsychological tests is reduced based on the number of SVTs failed up to two failures, after two (i.e., three or more), the decline is less significant with each SVT failed. Demographic variable have only a small effect on SVT performance with 1.2% of variance accounted for by education. Age was inversely correlated with SVT performance, indicating that, in this sample, the younger individuals failed more SVTs than did older individuals. There results although significant are likely an incidental finding. Our findings do show similar results with that of Green and colleagues (2001) in that about 50% of the variance in neuropsychological test performance can be accounted for by motivation as measured by internal SVTs.

Conflict of interest

None declared.

References

Andrews, F. M., Klem, L., Davidson, T. N., O'Malley, P. M., & Rodgers, W. L. (1981). *A guide for selecting statistical techniques for analyzing social science data* (2nd ed.). Institute for Social Research, University of Michigan.

Benton, A., Hamsher, K., Varney, N., & Spreen, O. (1983). *Contributions to neuropsychological assessment: A clinical manual*. Oxford University Press.

Brandt, J., Rubinsky, E., & Larson, G. (1985). Uncovering malingered amnesia. *Annals of the New York Academy of Science*, *44*, 502–503.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20*, 419–426.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dikmen, S. S., Machamer, J. E., Winn, H. R., & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, *9*, 80–90.

Green, P. (2003). *Word Memory Test*. Edmonton, AL: Green Publishing.

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*, 1045–1060.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*, 218–224.

Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J. E., Curtis, K. L., & Ord, J. S. (2009). The Reliable Digit Span Test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist*, *24*, 137–52.

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

Kozel, J. J., & Meyers, J. E. (1998). A cross-validation study of the Victoria Revision of the Category Test. *Archives of Clinical Neuropsychology*, *13*, 327–332.

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

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666–679.

Meyers, J. E. (2007). Malingering mild traumatic brain injury: Behavioral approaches used by both malingering actors and probable malingerers. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial: professional manual*. Odessa, FL: Psychological Assessment Resource.

Meyers, J. E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment*, *5*, 301–305.

Meyers, J. E., & Volbrecht, M. (1999). Detection of malingerers using the Rey Complex Figure and Recognition Trial. *Applied Neuropsychology*, *6*, 201–207.

Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, *18*, 261–276.

Meyers, J. E., & Rohling, M. L. (2004). Validation of the Meyers Short Battery on mild TBI patients. *Archives of Clinical Neuropsychology*, *19*, 637–651.

Meyers, J. E., Bayless, J., & Meyers, K. R. (1996). The Rey Complex Figure: Memory error patterns and functional abilities. *Applied Neuropsychology*, *3*, 89–92.

Meyers, J. E., Galinsky, A., & Volbrecht, M. (1999). Malingering and mild brain injury: How low is too low. *Applied Neuropsychology*, *6*, 208–216.

Meyers, J. E., Morrison, A. L., & Miller, J. C. (2001). How low is too low, Revisited: Sentence Repetition and AVLT-Recognition in the detection of malingering. *Applied Neuropsychology*, *8 (4)*, 234–241.

Meyers, J. E., Roberts, R. J., Bayless, J.D., Volkert, K.T., & Evitts, P.E. (2002). Dichotic listening: Expanded norms and clinical application. *Archives of Clinical Neuropsychology*, *17 (1)*, 79–90.

Reitan, R., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: Theory and interpretation*. Tucson: Neuropsychology Press.

Rohling, M. L., Meyers, J. E., & Millis, S. R. (2003). Neuropsychological impairment following traumatic brain injury: A dose–response analysis. *The Clinical Neuropsychologist*, *17*, 289–302.

Strauss, E., Sherman, E.M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Wechsler, D. (1997). *WAIS-III: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.