

DIALOGUE

# Symptom Validity Testing, Effort, and Neuropsychological Assessment



Erin D. Bigler<sup>1,2,3,4</sup>

<sup>1</sup>Department of Psychology, Brigham Young University, Provo, Utah

<sup>2</sup>Neuroscience Center, Brigham Young University, Provo, Utah

<sup>3</sup>Department of Psychiatry, University of Utah, Salt Lake City, Utah

<sup>4</sup>The Brain Institute of Utah, University of Utah, Salt Lake City, Utah

(RECEIVED November 14, 2011; FINAL REVISION February 4, 2012; ACCEPTED February 8, 2012)

## Abstract

Symptom validity testing (SVT) has become a major theme of contemporary neuropsychological research. However, many issues about the meaning and interpretation of SVT findings will require the best in research design and methods to more precisely characterize what SVT tasks measure and how SVT test findings are to be used in neuropsychological assessment. Major clinical and research issues are overviewed including the use of the “effort” term to connote validity of SVT performance, the use of cut-scores, the absence of lesion-localization studies in SVT research, neuropsychiatric status and SVT performance and the rigor of SVT research designs. Case studies that demonstrate critical issues involving SVT interpretation are presented. (*JINS*, 2012, 18, 632–642)

**Keywords:** Symptom validity testing, SVT, Effort, Response bias, Validity

Symptom validity testing (SVT) has emerged as a major theme of neuropsychological research and clinical practice. Neuropsychological assessment methods and procedures strive for and require the most valid and reliable techniques to assess cognitive and neurobehavioral functioning, to make neuropsychological inferences and diagnostic conclusions. From the beginnings of neuropsychology as a discipline, issues of test reliability and validity have always been a concern (Filskov & Boll, 1981; Lezak, 1976). However, neuropsychology’s initial focus was mostly on test development, standardization and the psychometric properties of a test and not independent measures of test validity. A variety of SVT methods are now available (Larrabee, 2007). While contemporary neuropsychological test development has begun to more directly incorporate SVT indicators embedded within the primary neuropsychological instrument (Bender, Martin Garcia, & Barr, 2010; Miller et al., 2011; Powell, Locke, Smigielski, & McCrea, 2011), traditional neuropsychological test construction and the vast majority of standardized tests currently in use do not. Current practice has

been to use what are referred to as “stand-alone” SVT measures that are separately administered during the neuropsychological examination (Sollman & Berry, 2011). SVT performance is then used to infer validity or lack thereof for the battery of all neuropsychological tests administered during that test session.

The growth in SVT research has been exponential. Using the search words “symptom validity testing” in a National Library of Medicine literature search yielded only one study before 1980, five articles during the 1980s, but hundreds thereafter. SVT research of the last decade has led to important practice conclusions as follows: (1) professional societies endorse SVT use (Bush et al., 2005; Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009), (2) passing a SVT infers valid performance, (3) SVT measures have good face validity as cognitive measures, all have components that are easy for healthy controls and even for the majority of neurological patients to pass with few or no errors, and (4) groups that perform below established cut-score levels on a SVT generally exhibit lower neuropsychological test scores. This last observation has been interpreted as demonstrating that SVT performance taps a dimension of effort to perform, where SVT “failure” reflects non-neurological factors that reduce neuropsychological test scores and invalidates findings

Correspondence and reprint requests to: Erin D. Bigler, Department of Psychology and Neuroscience Center, 1001 SWKT, Brigham Young University, Provo, UT 84602. E-mail: erin\_bigler@byu.edu

(West, Curtis, Greve, & Bianchini, 2011). On forced-choice SVTs, the statistical improbability of below chance performance implicates malingering (the examinee knows the correct answer but selects the incorrect to feign impairment).

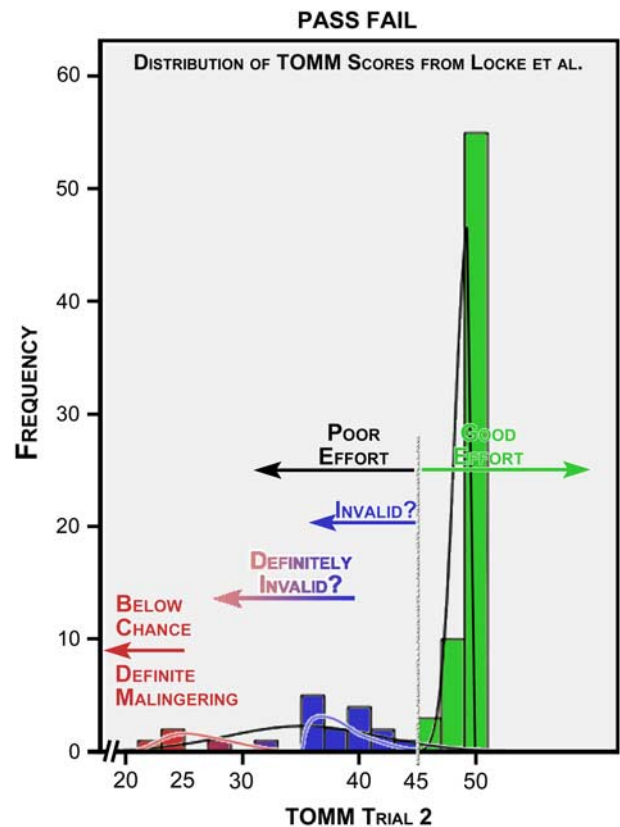
A quote from Millis (2009) captures the importance of why neuropsychology must address the validity of test performance:

All cognitive tests require that patients give their *best effort* (italics added) when completing them. Furthermore, cognitive tests do not directly measure cognition: they measure behavior from which we make inferences about cognition. People are able to consciously alter or modify their behavior, including their behavior when performing cognitive tests. Ostensibly poor or “impaired” test scores will be obtained if an examinee withholds effort (e.g., reacting slowly to reaction time tests). There are many reasons why people may fail to give best effort on cognitive testing: financial compensation for personal injury; disability payments; avoiding or escaping formal duty or responsibilities (e.g., prison, military, or public service, or family support payments or other financial obligations); or psychosocial reinforcement for assuming the sick role (Slick, Sherman, & Iverson, 1999). “... Clinical observation alone cannot reliably differentiate examinees giving best effort from those who are not.” (Millis & Volinsky, 2001, p. 2409)

This short review examines key SVT concepts and the “effort” term as used in neuropsychology. “Effort” seems to capture a clinical descriptive of patient test performance that, at first blush, seems straightforward enough. However, effort also has neurobiological underpinnings, a perspective often overlooked in SVT research and clinical application. Furthermore, does the term effort suggest intention, such as “genuine effort” – the patient is trying their best? Or if maximum “effort” is not being applied to test performance or exhibited by the patient, when does it reflect performance that may not be trustworthy? What is meant by effort?

### “EFFORT” – ITS MULTIPLE MEANINGS

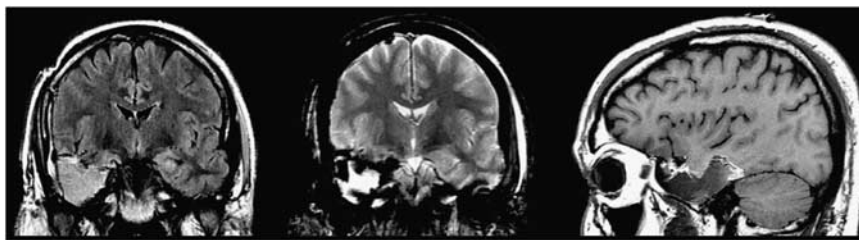
In the SVT literature when the “effort” term is linked with other nouns, verbs, and/or adjectives such statements appear to infer or explain a patient’s mental state, including motivation. For example it is common to see commentaries in the SVT literature indicating something about “cognitive effort,” “mental effort,” “insufficient effort,” “poor effort,” “invalid effort,” or even “faked effort.” Some of these terms suggest that neuropsychological techniques, and in particular the SVT measure itself, are capable of inferring intent? Can they (see discussion by Dressing, Widder, & Foerster, 2010)? There are additional terms often used to describe SVT findings including response bias, disingenuous, dissimulation, non-credible, malingered, or non- or sub-optimal, further complicating SVT nomenclature. There is no agreed upon consensus definition within neuropsychology of what effort means.



**Fig. 1.** The distribution of Test of Memory Malingerer (TOMM) scores that were above (green) the cut-point of 45 compared to those who performed below (blue-red). Note the bi-modal distribution of those who failed, compared to the peaked performance of 50/50 correct by those who pass. Below chance responding is the most accepted SVT feature indicative of malingering. As indicated by the shading of blue coloration emerging into red, as performance approaches chance, that likely reflects considerable likelihood of malingering. However, recalling that all of these patients had some form of ABI, those in blue, are much closer to the cut-point, raising the question of whether their neurological condition may contribute to their SVT performance.

### Below Cut-Score SVT Performance and Neuropsychological Test Findings: THE ISSUE

An exemplary SVT study has been done by Locke, Smigielski, Powell, and Stevens (2008). This study is singled out for this review because it was performed at an academic medical center (many SVT studies are based on practitioners’ clinical cases), had institutional review board (IRB) approval (most SVT studies do not), examined non-forensic (no case was in litigation although some cases had already been judged disabled and were receiving compensation), and was based on consecutive clinical referrals (independently diagnosed with some type of acquired brain injury [ABI] before the neuropsychological assessment) with all patients being seen for treatment recommendations and/or actual treatment. Figure 1 shows the distribution of pass/fail SVT scores where 21.8% performed below the SVT cut-point [pass  $\geq$  45/50 items correct on Trial 2 of the Test of Memory Malingerer (TOMM), Tombaugh, 1996], which Locke et al. defined as



**Fig. 2.** MRI showing partial temporal lobectomy. Pre-Surgery Test of Memory and Malinger (TOMM): Trial 1 = 50/50; Trial 2 = 50/50; Test of Neuropsychological Malinger (TNM) = 90% correct (see Hall and Pritchard, 1996). Post-Surgery: TOMM Trial 1 = 42/50; Trial 2 = 46/50; Delayed = 44/50; Rey 15-Item = 6/15; Word Memory Test (WMT): IR 67.5%, 30 min delay 75%; Free Recall 5%; Free Recall Delay = 7.5%; Free Recall Long Delay = 0.0.

constituting a group of ABI patients exhibiting “sub-optimal” effort. Of greater significance for neuropsychology is that the subjects in the failed SVT group, while being matched on all other demographic factors, performed worse and statistically lower across most (but not all) neuropsychological test measures.

As seen in Figure 1, the modal response is a perfect or near perfect TOMM score of 50/50 correct, presumably reflective of valid test performance. The fact that all subjects had ABI previously and independently diagnosed before the SVT being administered, and most performed without error, is a testament to the ease of performing an SVT. For those who scored below the cut-point and thereby “failed” the SVT measure, a distinct bi-modal distribution emerges. One group, shown in red (see Figure 1), hovers around chance (clearly an invalid performance) but the majority of others who fail, as shown in blue, hover much closer to, but obviously below, the SVT cut-point.

For the purposes of this review, the distinctly above-chance but below cut-score performing group will be labeled the “Near-Pass” SVT group. It is with this group where the clinician/researcher is confronted with Type I and II statistical errors when attempting to address validity issues of neuropsychological test findings. All current SVT methods acknowledge that certain neurological conditions may influence SVT performance, but few provide much in the way of guidelines as to how this problem should be addressed.

### Type II Error and the Patient With a “Near-Pass SVT” performance<sup>1</sup>

Two cases among many seen in our University-based neuropsychology program are representative of the problems with classifying neuropsychological test performance as invalid in the “Near-Pass” SVT performing patient. Figure 2 depicts the post-temporal lobectomy MRI in a patient with intractable epilepsy who underwent a partial right hippocampectomy. Pre-surgery he passes all SVT measures but post-surgery passes some but not others. Figure 3 shows abnormal medial temporal damage in a patient with herpes simplex encephalitis

who fails SVT measures. In both cases, SVT failure is not below chance, both patients have distinct, bona fide and unequivocal structural damage to critical brain regions involved in memory. Is not their SVT performance a reflection of the underlying damage and its disruption of memory performance? If a neuropsychologist interpreted these test data as invalid because of “failed” SVT performance, is that not ignoring the obvious and committing a Type II error? So when is “failed” SVT performance just a reflection of underlying neuropathology? For example, cognitive impairment associated with Alzheimer’s disease is sufficient to impair SVT performance resulting in below recommended cut-score levels and therefore constituting



**Fig. 3.** The fluid attenuated inversion recovery (FLAIR) sequence magnetic resonance imaging (MRI) was performed in the sub-acute stage demonstrating marked involvement of the right (arrow) medial temporal lobe region of this patient (see arrow). The patient was attempting to return to college and was being evaluated for special assistance placement. On the Wechsler Memory Scale (WMS-IV) he obtained the following: WMS-IV: Audio Memory = 87; Visual Memory = 87; Visual Word Memory = 73; Immediate Memory = 86; Delayed Memory = 84. Immediate Recall: 77.5% (Fail); Delayed Recall: 72.5% (Fail); Consistency: 65.0% (Fail); Multiple Choice: 50.0% (Warning); Paired Associate: 50.0% (Warning); Free Recall: 47.5%; Test of Memory and Malinger (TOMM): Trial 1: 39, Trial 2: 47.

<sup>1</sup> It is impossible to discuss SVT measures without discussing some by name. This should not be considered as any type of endorsement or critique of the SVT mentioned. SVT inclusion of named tests in this review occurred merely on the basis of the research being cited.



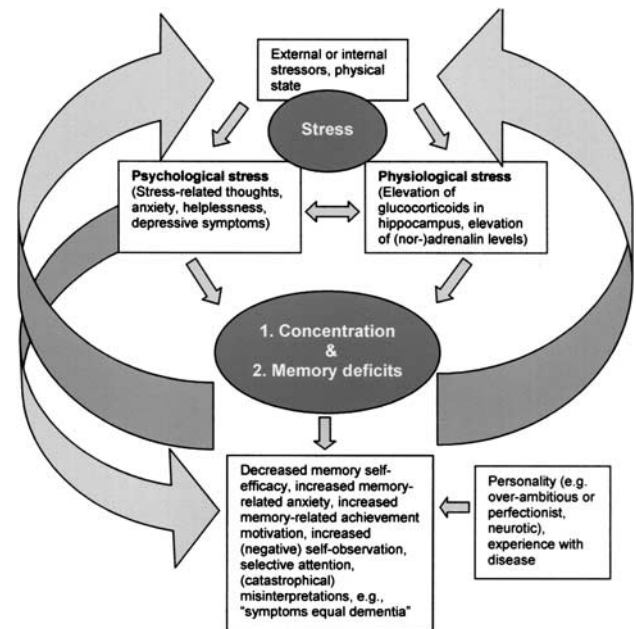
a SVT failure (Merten, Bossink, & Schmand, 2007). However, in such a circumstance impaired neuropsychological performance and the low SVT score are both thought to reflect genuinely impaired cognitive ability.

### No Systematic Study Lesion/Localization Studies of SVT performance

Neuropsychological assessment has a long tradition of examining lesion effects (or lack thereof) on neuropsychological test performance (Cipolotti & Warrington, 1995). There are no systematic studies of lesion effects on SVT performance. Despite the assumption of minimal requirements to perform a SVT task, episodic memory is being tapped where functional neuroimaging studies demonstrate expected medial temporal, cingulate and parietofrontal attentional network activation during SVT performance (Allen, Bigler, Larsen, Goodrich-Hunsaker, & Hopkins, 2007; Browndyke et al., 2008; Larsen, Allen, Bigler, Goodrich-Hunsaker, & Hopkins, 2010). Goodrich-Hunsaker and Hopkins (2009) and Wu, Allen, Goodrich-Hunsaker, Hopkins, & Bigler (2010) in case studies have shown that five patients with hippocampal damage (3 with anoxic injury, 2 with TBI) can pass SVTs. However, this woefully under samples the possible lesions and abnormalities that have the potential to directly disrupt SVT performance. While not a lesion study, Merten et al. (2007) have demonstrated SVT failure related to dementia severity and Gorissen, Sanz, and Schmand (2005) demonstrated high SVT failure rates in neuropsychiatric patients, particularly those with schizophrenia. If certain neuropathological conditions are more likely to affect SVT performance, such findings would be critical for SVT interpretation. Currently, there are no recommended adjustments in SVT cut-scores based on location, size or type of lesion that may be present.

### Illness Behavior and Cognitive Performance

In non-litigating neurological and neuropsychiatric clinical populations, 15 to 30% or higher SVT failure rates have been reported (Williams, 2011). Does this mean invalid neuropsychological test data occurs in up to one-third of all patients seen for neuropsychological assessment based on not passing an externally administered SVT? Are these patients malingering? What is the role of SVT performance in rendering differential diagnoses involving malingering, somatoform disorder and other neuropsychiatric conditions? Figure 4 from Metternich, Schmidtke, and Hull (2009) provides a model showing the overlap between functional and constitutional memory factors that may adversely affect cognition in the neuropsychiatric patient. In reviewing this model, the reader should note that there are numerous meta-cognitive as well as neural pathways that could legitimately disrupt SVT performance as a result of the disorder. Wager-Smith and Markou (2011) review the effects of stress, cytokine, and neuroinflammatory reactions that relate to “sickness behavior” and impaired cognition. In the context of the Metternich et al. (2009) model, sickness behaviors may interact with stress mediated biological factors that



**Fig. 4.** Theoretical model postulating how stress may influence cognition. Primary influences may come from purely psychosocial variables, directly from physiologically activated stress variables or the combination of the two. Reprinted from *Journal of Psychosomatic Research*, Volume 66, Issue 05. “How are memory complaints in functional memory disorder related to measures of affect, metamemory and cognition?” by Birgitta Metternich, Kalasus Schmidtke and Michael Hull, pp. 435–444. Copyright © 2009 with permission from Elsevier.

appear psychological yet disrupt cognition. Does any of this reflect differences in how SVT cut scores should be established if a known neuropsychiatric disorder is present?

### Diagnosis Threat and SVT Performance

The “diagnosis threat” literature clearly demonstrates both experimentally and clinically that performance *expectations* influence actual cognitive test performance and the perceived influence of symptoms on cognitive test results (Ozen & Fernandes, 2011; Suhr & Gunstad, 2005). Likewise, placebo research on cognitive performance plainly demonstrates the influential role that expectations have on symptom generation and test performance (Pollo & Benedetti, 2009). Thus, psychological state and trait characteristics and the perception of well-being *versus* “illness” may influence cognitive performance (Pressman & Cohen, 2005; Walkenhorst & Crowe, 2009). Is “near-pass” SVT performance an expected human dimension of diagnosis threat? Can these factors be disentangled from the degree and type of brain injury, medication status, litigation status, levels of psychological distress including premorbid conditions and other non-neurological factors by SVT performance (see discussion by Suhr, Tranel, Wefel, & Barrash, 1997)?

### Effort or Ability?

If SVT performance required minimal to no cognitive effort, then experimental paradigms using cognitive load as a

distraction, should result in minimal to little change in SVT performance. Batt, Shores, and Chekaluk (2008) examined non-litigating severe TBI patients on SVT measures during a task where distraction occurred during the SVT learning phase, demonstrating the influence of cognitive processing on SVT performance. Cognitive neuroscience uses simple cognitive tasks, as simple as SVT measures, to experimentally manipulate conditions to tease out neural and experimental effects on cognition (Graham, Barense, & Lee, 2010). Unfortunately, other than the Batt et al. and a handful of other studies, the cognitive neuroscience of SVT performance has been ignored. For example, it is unknown whether the foil stimuli used in an SVT task are equivalent or not, or are uniquely influenced by certain types of structural damage, or neuropsychiatric or neurological condition.

### Design Issues in SVT Research

The review to this point has raised several interpretative questions that occur in Near-Pass SVT subjects. The answer to these questions requires better designed SVT studies that address ambiguities of past SVT findings. Williams (2011) points out the necessity of some “messy” SVT research designs due to the impossibility of getting genuine malingerers to volunteer for standardization studies. Standardization studies have had to rely on simulator studies and clinical samples; mostly ones of convenience and mostly forensic samples.

#### *Circular reasoning, tautology and SVT research*

If one uses SVT performance as the only index of effort and then concludes that SVT failure is a sign of “poor effort,” yet there are no other independent measures of what may be test behavior compliance or willingness to engage, process, and perform, or even malingering, then is this not a tautological argument? In such studies the only classifier defining poor effort is the SVT performance itself. While such studies often classify subjects by secondary gain identifiers (litigation, disability determination, etc.), such identifiers are not *direct* measures of effort, only of secondary gain.

Tautology also involves the unnecessary repetitive use of different words with the same meaning. The list of terms being used interchangeably with SVT and effort include response bias, invalid or failed performance, symptom amplification, performance exaggeration, underperformance or distortion, symptom embellishment, disingenuous, sub-optimal, poor effort, non-credible, faked and malingered. It is not uncommon to see statements like – “Failed SVT performance was associated with *invalid* neuropsychological test performance that was deemed *non-credible* due to *sub-optimal effort*.” The tautological problems with such a statement should be obvious.

### Rigor of SVT studies

Class I and II level research as endorsed by the National Institutes of Health (NIH) and all major medical societies involves independently conducted investigations that have some external

review and monitoring where investigators are independent of the outcome at *all* levels of the investigation (Edlund, Gronseth, So, & Franklin, 2004). The best Class I and II investigations are those where *a priori* consensus diagnostic standards are in place that are independent of any outcome measure, data collected prospectively and independently from those in charge of their analysis, where clinicians involved in diagnostic decision making are independent of those who analyze the data, and who are also blinded. In Class I or II investigations, clinicians and data managers cannot also be the statisticians. Explicitly different roles at all levels of data acquisition, tabulation, analysis, and report writing increases the likelihood of unbiased findings. Institutional based investigations require human subjects review, consent, and IRB approval.

Few current SVT studies meet Class I or II level research or are subject to IRB approval. SVT studies that come from clinical practitioners in private practice not affiliated with an institution do not fall under any external review process whatsoever. Important investigational research comes from clinicians in private practice but rarely does this research meet a Class I or II standard.

### IS THERE A NEUROBIOLOGY OF DRIVE, EFFORT, MOTIVATION, AND ATTENTION?

Much of the discussion up to this point has focused on cognitive elements of SVT performance but there is also a behavioral dimension that centers on the neurobiology of drive, effort, and motivation (Sarter, Gehring, & Kozak, 2006). Patients with frontotemporolimbic damage may be apathetic with problems sustaining drive and goal-directed behaviors (Lezak, Howieson & Loring, 2004). Apathy is a common consequence of traumatic brain injury (TBI; Marin & Wilkosz, 2005). What happens during neuropsychological assessment of the patient with neurogenic drive and motivation problems? The following test scores were obtained in a patient approximately one year post-TBI who the family described as unmotivated; neuroimaging demonstrated extensive bi-frontal and right parietal encephalomalacia and generalized atrophy: TOMM: Trial 1 = 45/50; Trial 2 = 50/50; Rey 15-Item: 10/15; Word Memory Test (WMT): IR (immediate recognition) = 78%; DR (delayed recognition) = 85%; CNS (consistency response) = 78%. The 45/50 on TOMM Trial 1 represents a pass, but is right at the cut-score with a perfect 50/50 on Trial 2 representing a pass by all standards. The Rey 15-Item performance represents a pass (although a borderline score by some standards); however, a 78% correct on the WMT IR and CNS Scales represents a “failure” by WMT standards. How does brain damage to motivational and attentional networks affect SVT performance and should patients with obvious structural lesions be evaluated by different cut-score standards?

### Which Test to Use?

There are now numerous SVT measures available for use in general neuropsychological practice (Grant & Adams, 2009;

Lezak, Howieson, & Loring, 2004; Mitrushina, Boone, Razani, & D'Elia, 2005; Strauss, Sherman, & Spreen, 2006; Tate, 2010) and potentially even more in a forensic setting (Boone, 2007; Larrabee, 2007; Morgan & Sweet, 2009). While SVT use is endorsed by professionals, SVT test selection relies solely on the judgment of the researcher and/or clinician. With such a broad array of SVT measures that can be used, which ones *should* be used and in what circumstances? Also, a reasonable argument has been made that multiple SVTs are needed, especially in any lengthy or forensic assessment (Boone, 2009; Larrabee, 2008), but again no agreed upon professional standards as to the correct number, in what order, and in what context. Administration of multiple SVT measures also raises other questions when failures on some but not others occur and whether there is an order effect in SVT test administration (Ryan, Glass, Hinds, & Brown, 2010)?

## THE PROBLEM WITH CUT SCORES

Dwyer (1996) reviewed the methods of cut-score development concluding that cut-scores (a) always entail judgment; (b) inherently result in some misclassification, (c) impose artificial "pass/fail" dichotomies and (d) no "true" cut scores exist (p. 360). Given Dwyer's comments, should cut scores be adjusted or customized to specific clinical conditions? Kirkwood and Kirk (2010) in a university-based assessment clinic evaluated 193 consecutively referred children with mild TBI. This study had IRB approval and these investigators examined both the TOMM and Medical Symptom Validity Test (MSVT; Green, 2004). Using their terminology, Kirkwood and Kirk (2010) found 17% of the sample to exhibit "suboptimal effort." Only one failure was thought to be influenced by litigation. They also attempted to identify other potential sources for SVT failure, including impulsivity, distractibility during testing, pre-injury neuropsychiatric diagnosis and potential effects of reading disability. This study unmistakably demonstrates the complexities and potential issues in a clinical sample that may lead to sub-optimal performance.

In the Merten et al. (2007) study mentioned earlier that examined bona fide neurological patients with and without clinically obvious symptoms only 15/48 (31%) passed all SVTs. Only 1/24 (4%) of those with clinically obvious cognitive symptoms was able to pass all SVTs. These authors conclude that the "... Results clearly show that many of these bona fide patients fail on most SVTs. Had the recommended cutoffs been applied rigorously without taking the clinical picture into consideration, these patients would have been incorrectly classified as exerting insufficient effort." (p. 314). Donders and Strong (2011) attempted to replicate a logistic regression method that had been developed by Wolfe et al. (2010) to identify embedded effort indicators on the California Verbal Learning Test (CVLT). They not only applied the logistic regression method of Wolfe et al. but also used an externally administered SVT. However, the limits of interrelationship and inter-test agreement between actual test performances, embedded indicators of effort and the external SVT led them to conclude that this method was not ready for clinical application.

These studies simply underscore the difficulties of what needs to be addressed and accounted for in the neuropsychological application of current SVT technology. Similarly, Powell et al. (2011) show that supposed markers of suboptimal effort using the Trail Making Test, also have limited predictive ability; again demonstrating the problem of disentangling true neuropsychological performance from associated elements of effort, drive, motivation, and attention when attempting to use embedded methods that were not explicitly designed to simultaneously assess validity.

Cut-scores are a necessary part of contemporary neuropsychological testing as they provide a method for classification but cut-scores are best used in the context of guidelines, rather than a dichotomous defining point for presence or absence of a deficit (Strauss et al., 2006).

## False Memory and Dissociative Reactions

Mental health issues as they relate to "false memory" have been the topic of considerable controversy in clinical, research, and legal settings (Loftus & Davis, 2006). Interestingly, even an animal model of false memory has been developed (McTighe, Cowell, Winters, Bussey, & Saksida, 2010). Are some SVT failures by neurological or neuropsychiatric patients generated by "false memories"? For example, cognitive neuroscience often examines how confident a subject is in their response, when assessing false memory (Moritz & Woodward, 2006). No SVT studies to date have tackled this dimension.

## Does Failed SVT Always Equate With Invalid Performance for All Neuropsychological Measures?

In the Locke et al. (2008) investigation not all neuropsychological test scores were significantly suppressed in the failed SVT group. The Category Test and two of the three Wisconsin Card Sorting measures did not differ between the group that "passed" the SVT and the group that "failed" nor did scores on the Beck Depression and Anxiety scales. Whitney, Shepard, Mariner, Mossbarger, and Herman (2010) found that Wechsler Test of Adult Reading (WTAR) scores were no different between those with passed or failed SVT performance suggesting that WTAR findings remain "robust even in the face of suboptimal effort" (p. 196). Does this mean valid neuropsychological test findings occur with some tests even in the presence of SVT "failure"?

## SVT Deception

Is the SVT measure infallible? DenBoer and Hall (2007) have shown that simulators can be "taught" how to detect SVT tasks and pass them and go on to fail the more formal neuropsychological measures (see also Rüsseler, Brett, Klaue, Sailer, & Munte, 2008). If the SVT task can be faked how would the clinician and/or researcher know?

## CONCLUSIONS

SVT findings may offer important information about neuropsychological test performance, but if an oversimplified view



of SVT-test behavior dichotomizes neuropsychological performance as either valid (above cut-score) or invalid (below cut-score), clinically important Type I and II errors are unavoidable. As shown in this review, patients with legitimate neurological and/or neuropsychiatric conditions fail SVTs for likely neurogenic factors. There can be no debate that issues of poor effort and secondary gain may have such a profound effect as to completely invalidate neuropsychological test findings rendering them uninterruptable (see Stevens, Friedel, Mehren, & Merten, 2008). However, considerably more SVT research is needed to address the issues raised in this review.

## ACKNOWLEDGMENTS

Parts of this were presented as a debate at the 38th Annual International Neuropsychological Society meetings on “Admissibility and Appropriate Use of Symptom Validity Science in Forensic Consulting” in Acapulco, Mexico moderated by Paul M. Kaufmann, J.D., Ph.D. Dr. Bigler co-directs Brigham Young University’s Neuropsychological Assessment and Research Clinic that provides a service to the community evaluating a broad spectrum of clinical referrals. For legal referrals the University is compensated for the evaluation. Dr. Bigler also performs forensic consultation for which he is directly compensated. His research is National Institutes of Health (NIH) funded, but no NIH grant funds directly supported the writing of this commentary. Dr. Bigler has no financial interest in any commercial symptom validity test (SVT) measure. The assistance of Jo Ann Petrie, Thomas J. Farrer, and Tracy J. Abildskov in the preparation of this manuscript is gratefully acknowledged.

## REFERENCES

- Allen, M.D., Bigler, E.D., Larsen, J., Goodrich-Hunsaker, N.J., & Hopkins, R.O. (2007). Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives. *Brain Injury*, *21*, 1425–1428. doi:10.1080/02699050701769819
- Batt, K., Shores, E.A., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society*, *14*, 1074–1080. doi:10.1017/S135561770808137X
- Bender, H.A., Martin Garcia, A., & Barr, W.B. (2010). An interdisciplinary approach to neuropsychological test construction: Perspectives from translation studies. *Journal of the International Neuropsychological Society*, *16*, 227–232. doi:10.1017/S1355617709991378
- Boone, K.B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: The Guilford Press.
- Boone, K.B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*, 729–741. doi:10.1080/13854040802427803
- Browndyke, J.N., Paskavitz, J., Sweet, L.H., Cohen, R.A., Tucker, K.A., Welsh-Bohmer, K.A., ... Schmechel, D.E. (2008). Neuroanatomical correlates of malingered memory impairment: Event-related fMRI of deception on a recognition memory task. *Brain Injury*, *22*, 481–489. doi:10.1080/02699050802084894
- Bush, S.S., Ruff, R.M., Troster, A.I., Barth, J.T., Koffler, S.P., Pliskin, N.H., ... Silver, C.H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*, 419–426. doi:10.1016/j.acn.2005.02.002
- Cipolotti, L., & Warrington, E.K. (1995). Neuropsychological assessment. *Journal of Neurology, Neurosurgery and Psychiatry*, *58*, 655–664.
- DenBoer, J.W., & Hall, S. (2007). Neuropsychological test performance of successful brain injury simulators. *The Clinical Neuropsychologist*, *21*, 943–955. doi:10.1080/13854040601020783
- Donders, J., & Strong, C.A. (2011). Embedded effort indicators on the California Verbal Learning Test–Second Edition (CVLT-II): An attempted cross-validation. *The Clinical Neuropsychologist*, *25*, 173–184. doi:10.1080/13854046.2010.536781
- Dressing, H., Widder, B., & Foerster, K. (2010). [Symptom validity tests in psychiatric assessment: A critical review]. *Versicherungsmedizin/herausgegeben von Verband der Lebensversicherungs-Unternehmen e.V. und Verband der Privaten Krankenversicherung e.V.*, *62*, 163–167.
- Dwyer, C.A. (1996). Cut scores and testing: Statistics, judgment, truth, and error. *Psychological Assessment*, *8*, 360–362.
- Edlund, W., Gronseth, G., So, Y., & Franklin, G. (2004). *Clinical practice guideline process manual: For the Quality Standards Subcommittee (QSS) and the Therapeutics and Technology Assessment Subcommittee (TTA)*. St. Paul: American Academy of Neurology.
- Filskov, S.B., & Boll, T.J. (1981). *Handbook of clinical neuropsychology*. New York: John Wiley & Sons.
- Goodrich-Hunsaker, N.J., & Hopkins, R.O. (2009). Word memory test performance in amnesic patients with hippocampal damage. *Neuropsychology*, *23*, 529–534. doi:10.1037/a0015444
- Gorissen, M., Sanz, J.C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, *78*, 199–208. doi:10.1016/j.schres.2005.02.016
- Graham, K.S., Barense, M.D., & Lee, A.C. (2010). Going beyond LTM in the MTL: A synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia*, *48*, 831–853. doi:10.1016/j.neuropsychologia.2010.01.001
- Grant, I., & Adams, K.M. (2009). *Neuropsychological assessment of neuropsychiatric and neuromedical disorders*. New York: Oxford University Press.
- Green, P. (2004). *Medical Symptom Validity Test (MSVT) for Microsoft Windows: User’s Manual*. Edmonton, Canada: Green’s Publishing.
- Hall, H., & Pritchard, D. (1996). *Detecting malingering and deception: Forensic distortion analysis*. Florida: St. Lucie Press.
- Heilbronner, R.L., Sweet, J.J., Morgan, J.E., Larrabee, G.J., & Millis, S.R. (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. doi:10.1080/13854040903155063
- Kirkwood, M.W., & Kirk, J.W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist*, *24*, 860–872. doi:10.1080/13854040903527287
- Larrabee, G.J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.
- Larrabee, G.J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666–679. doi:10.1080/13854040701494987

- Larsen, J.D., Allen, M.D., Bigler, E.D., Goodrich-Hunsaker, N.J., & Hopkins, R.O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury*, *24*, 89–99. doi:10.3109/02699050903508218
- Lezak, M.D. (1976). *Neuropsychological assessment*. New York: Oxford University Press.
- Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological assessment*. New York: Oxford.
- Locke, D.E., Smigielski, J.S., Powell, M.R., & Stevens, S.R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *Neurorehabilitation*, *23*, 273–281.
- Loftus, E.F., & Davis, D. (2006). Recovered memories. *Annual Review of Clinical Psychology*, *2*, 469–498. doi:10.1146/annurev.clinpsy.2.022305.095315
- Marin, R.S., & Wilkosz, P.A. (2005). Disorders of diminished motivation. *The Journal of Head Trauma Rehabilitation*, *20*, 377–388.
- McTighe, S.M., Cowell, R.A., Winters, B.D., Bussey, T.J., & Saksida, L.M. (2010). Paradoxical false memory for objects after brain damage. *Science*, *330*, 1408–1410. doi:10.1126/science.1194780
- Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, *29*, 308–318. doi:10.1080/13803390600693607
- Metternich, B., Schmidtke, K., & Hull, M. (2009). How are memory complaints in functional memory disorder related to measures of affect, metamemory and cognition? *Journal of Psychosomatic Research*, *66*, 435–444. doi:10.1016/j.jpsychores.2008.07.005
- Miller, J.B., Millis, S.R., Rapport, L.J., Bashem, J.R., Hanks, R.A., & Axelrod, B.N. (2011). Detection of insufficient effort using the advanced clinical solutions for the Wechsler Memory Scale, Fourth edition. *The Clinical Neuropsychologist*, *25*, 160–172. doi:10.1080/13854046.2010.533197
- Millis, S.R. (2009). Methodological challenges in assessment of cognition following mild head injury: Response to Malojcic et al. 2008. *Journal of Neurotrauma*, *26*, 2409–2410. doi:10.1089/neu.2008.0530
- Millis, S.R., & Volinsky, C.T. (2001). Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology*, *23*, 809–828.
- Mitrushina, M., Boone, K.B., Razani, J., & D'Elia, L.F. (2005). *Handbook of normative data for neuropsychological assessment*. New York: Oxford.
- Morgan, J.E., & Sweet, J.J. (2009). *Neuropsychology of malingering casebook*. New York: Psychology Press.
- Moritz, S., & Woodward, T.S. (2006). Metacognitive control over false memories: A key determinant of delusional thinking. *Current Psychiatry Reports*, *8*, 184–190.
- Ozen, L.J., & Fernandes, M.A. (2011). Effects of “diagnosis threat” on cognitive and affective functioning long after mild head injury. *Journal of the International Neuropsychological Society*, *17*, 219–229. doi:10.1017/S135561771000144X
- Pollo, A., & Benedetti, F. (2009). The placebo response: Neurobiological and clinical issues of neurological relevance. *Progress in Brain Research*, *175*, 283–294. doi:10.1016/S0079-6123(09)17520-9
- Powell, M.R., Locke, D.E., Smigielski, J.S., & McCrea, M. (2011). Estimating the diagnostic value of the Trail Making Test for suboptimal effort in acquired brain injury rehabilitation patients. *The Clinical Neuropsychologist*, *25*, 108–118. doi:10.1080/13854046.2010.532912
- Pressman, S.D., & Cohen, S. (2005). Does positive affect influence health? *Psychological Bulletin*, *131*, 925–971. doi:10.1037/0033-2909.131.6.925
- Rüsseler, J., Brett, A., Klaue, U., Sailer, M., & Munte, T.F. (2008). The effect of coaching on the simulated malingering of memory impairment. *BMC Neurology*, *8*, 37. doi:10.1186/1471-2377-8-37
- Ryan, J.J., Glass, L.A., Hinds, R.M., & Brown, C.N. (2010). Administration order effects on the Test of Memory Malingering. *Applied Neuropsychology*, *17*, 246–250. doi:10.1080/09084282.2010.499802
- Sarter, M., Gehring, W.J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, *51*, 145–160. doi:10.1016/j.brainresrev.2005.11.002
- Slick, D.J., Sherman, E.M., & Iverson, G.L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.
- Sollman, M.J., & Berry, D.T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, *26*, 744–789.
- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191–200. doi:10.1016/j.psychres.2007.01.003
- Strauss, E., Sherman, E.M.S., & Spreen, O. (2006). *A compendium of neuropsychological tests*. New York: Oxford University Press.
- Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, *19*, 500–514.
- Suhr, J.A., & Gunstad, J. (2005). Further exploration of the effect of “diagnosis threat” on cognitive performance in individuals with mild head injury. *Journal of the International Neuropsychological Society*, *11*, 23–29. doi:10.1017/S1355617705050010
- Tate, R.L. (2010). *A compendium of tests, scales and questionnaires: The practitioner's guide to measuring outcomes after acquired brain impairment*. New York: Psychology Press.
- Tombaugh, T. (1996). *TOMM: Test of memory malingering*. New York: Multi-Health Systems.
- Wager-Smith, K., & Markou, A. (2011). Depression: A repair response to stress-induced neuronal microdamage that can grade into a chronic neuroinflammatory condition? *Neuroscience and Biobehavioral Reviews*, *35*, 742–764. doi:10.1016/j.neubiorev.2010.09.010
- Walkenhorst, E., & Crowe, S.F. (2009). The effect of state worry and trait anxiety on working memory processes in a normal sample. *Anxiety, Stress, and Coping*, *22*, 167–187. doi:10.1080/10615800801998914
- West, L.K., Curtis, K.L., Greve, K.W., & Bianchini, K.J. (2011). Memory in traumatic brain injury: The effects of injury severity and effort on the Wechsler Memory Scale-III. *Journal of Neuropsychology*, *5*, 114–125. doi:10.1348/174866410X521434
- Whitney, K.A., Shepard, P.H., Mariner, J., Mossbarger, B., & Herman, S.M. (2010). Validity of the Wechsler Test of Adult Reading (WTAR): Effort considered in a clinical sample of U.S. military veterans. *Applied Neuropsychology*, *17*, 196–204. doi:10.1080/09084282.2010.499787
- Williams, J.M. (2011). The malingering factor. *Archives of Clinical Neuropsychology*, *26*, 280–285. doi:10.1093/arclin/acr009



Wolfe, P.L., Millis, S.R., Hanks, R., Fichtenberg, N., Larrabee, G.J., & Sweet, J.J. (2010). Effort indicators within the California Verbal Learning Test-II (CVLT-II). *The Clinical Neuropsychologist*, *24*, 153–168. doi:10.1080/13854040903107791

Wu, T.C., Allen, M.D., Goodrich-Hunsaker, N.J., Hopkins, R.O., & Bigler, E.D. (2010). Functional neuroimaging of symptom validity testing in traumatic brain injury. *Psychological Injury & Law*, *3*, 50–62. doi:DOI 10.1007/s12207-010-9067-y

doi:10.1017/S1355617712000409

## DIALOGUE RESPONSE

# Response to Larrabee

Erin D. Bigler

Neuropsychology needs objective methods that confidently and accurately reflect the validity of brain-behavior relationships as measured by neuropsychological assessment techniques. Symptom validity testing (SVT) has emerged as a method designed to address validity of neuropsychological test performance; but, just like the field of neuropsychology, SVT research is new and evolving. Within any new research endeavor, first generation studies often demonstrate broad support for a new construct but as the research expands more complex issues arise that require refinements in theory and practice (Oner & Dhert, 2011). Such is the case with SVT research and its clinical application. One goal of the dialogue with Larrabee on the current status of SVT research and clinical application was to highlight areas of agreement and disagreement. My review challenges some SVT assumptions, pointing out the need for refinements in methods and theory, calling for improved research designs that will hopefully lead to a more complete understanding of SVT use and interpretation in neuropsychological assessment.

Larrabee (this issue), in response to my SVT review (see Bigler, this issue), argues for a change in terminology, abandoning the singular term “effort” in favor of “performance validity” and “symptom validity” and offers cogent reasoning and research to support such a distinction. In my opinion, the term effort as a singular descriptor in neuropsychology should be abandoned in favor of the performance validity and symptom validity terms as suggested by Larrabee in his commentary. As already stated in the critique there are simply too many potential meanings suggested with just the term effort or “effort tests,” spanning the biological to inferring intent. From the biological, effort suggests neural factors associated with basic drives and emotional states (see Sarter, Gehring, & Kozak, 2006). Within cognitive neuroscience, effort relates directly to complexity of stimulus processing (Kohl, Wylie, Genova, Hillary, & Deluca, 2009) and levels of motivation (Bonnefond, Doignon-Camus, Hoefl, & Dufour, 2011; Harsay et al., 2011). In forensic and

applied neuropsychology, the effort term suggests some intention on the subject’s part where poor effort may be equated with malingering (see Williams, 2011). These multiple meanings make the term imprecise when used in neuropsychological parlance to describe test behavior. The “performance validity” and “symptom validity” terminology represent far more accurate descriptors of what is being assessed and neuropsychology will be better served by following Larrabee’s recommendation.

There are also two basic agreements on what may be considered SVT tenets: (1) questions of “symptom” and “performance” invalidity are proportional to the number SVT items *not* passed and, (2) close to or below chance SVT test performance levels are the clearest and most indisputable indicators for invalidity. In my opinion, little debate about the above two points is needed. For forced-choice SVT measures, invalid neuropsychological test performance may be assumed as SVT performance falls substantially below a conventionally established cut-score. SVT performance at, near, or below chance reflects invalid test performance.

Despite these points of agreement, two major SVT topics where our opinions diverge are: (1) the “false positive/false negative problem and interpretative validity issues” and, (2) the “rigor” of SVT study designs.

## THE FALSE POSITIVE PROBLEM AND INTERPRETATIVE VALIDITY ISSUES

The most effective SVT will minimize false positive and negative classifications with the false positive typically being the more serious error. False positive classification occurs when failed SVT scores are used to designate invalid neuropsychological test performance when in fact, the “failed” SVT performance occurs *because* of the underlying neurological and/or neuropsychiatric condition. The clinical gravity of a false positive SVT decision for neuropsychology is obvious—in the face of a false positive SVT indicating