Use of Symptom Validity Tests and Performance Validity Tests in Disability Determinations[1]

Erin D. Bigler, Ph.D., ABPP

Brigham Young University, Provo

Disclaimer: The authors are responsible for the content of this article, which does not necessarily represent the views of the Institute of Medicine

INTRODUCTION

Certain neurological conditions produce motor and sensory deficits that are unmistakable, clinically evident, and objective to identify. Other neurological conditions have laboratory, radiological, or neuroimaging biomarkers that are definitive for the condition being examined. Likewise, over time progressive neurological disorders like Alzheimer's have an array of distinct indicators and laboratory tests that provide validity and assurance that the tests and examination methods objectify the level of impairment and disability. The above mentioned conditions require no additional proof that a valid observation has been made.

How mental health and neuropsychiatric disorders are diagnosed and disability identified and quantified represents a different, less objective clinical picture, especially when it comes to psychological and neuropsychological test results. Whether it be some feature of cognition, filling out a questionnaire or symptom survey, or measuring some behavioral characteristic, all aspects of psychological assessment include an element of subjectivity. All require compliance on the part of the examinee performing the test. If the examinee is not putting forth some reasonable level of effort to perform, test scores may not accurately reflect the examinee's true ability level. If the test results are to be used to establish a diagnosis, treatment

[1]The writing of this review was commissioned by the Institute of Medicine. The author provides forensic consultation in the course of his clinical work as co-director of the Neuropsychological Research and Clinical Assessment Laboratory at Brigham Young University.

recommendations, and/or a disability determination, invalid test results could be misleading and even represent an attempt on the part of the examinee to falsify the examination process.

How to identify falsified impairments, malingering, and exaggerated deficits is a key topic in healthcare. Identification of valid psychological and neuropsychological test findings represents a major societal, insurance, and government issue as test results from these kinds of measures are used in disability determination. Obviously, the Social Security Administration (SSA) should only grant disability status to those with legitimate claims. Dissimulation studies using psychological assessment techniques among Social Security disability income claimants have been conducted for more than 30 years implicating substantial rates of dissimulation (Griffin et al., 1996), resulting in substantial financial costs (Chafetz and Underhill, 2013). How does one identify non-credible claims using psychological and neuropsychological tests? What tests should be used and how should they be interpreted?

A myriad of psychological, cognitive, and behavioral assessment methods have been developed to assess a wide spectrum of conditions related to mental and neurological health (Lezak et al., 2012). However, most of these measures have no or only limited built-in validity indicators, as the tradition in psychological test development assumed cooperation, sufficient effort, and valid test results would naturally occur within the one-on-one testing setting. Since intrinsic or internal measures of valid examinee performance for individual psychological/neuropsychological tests were absent from the original tests, external measures were developed and purportedly were designed to be general validity indicators of all other tests administered during a testing session. In the beginning these external measures became known as symptom validity tests (SVTs) or effort tests (ETs). During the last decade, the term "performance validity test" (PVT) was introduced (see Bigler, 2012; Larrabee, 2012; Van Dyke et al., 2013). PVT is probably the better term to use when describing validity for measures that assess cognitive, motor, sensory, or some behavioral task or ability that requires actual performance of the task. Validity of symptom reporting or endorsement where no task is performed other than reporting of the symptom or completing a questionnaire is probably best captured by the term SVT. Despite these distinctions, SVT, PVT, and ET are often used interchangeably. As will be discussed, there are many complexities that surround the term "effort" and therefore use of the term ET is discouraged. Because many publications do not distinguish between SVTs and PVTs, the combined PVT/SVT acronym is used frequently in this

review. For the most part, individual PVT/SVT tests will not be mentioned, and the review focuses mainly on conceptual aspects of these procedures. Furthermore, there is no large-scale comparative study across different SVT/PVT tasks that has measured their effectiveness or established a standard.

Without debate, the main finding from three decades of SVT/PVT research is that how well an examinee performs on a SVT/PVT relates to performance on other tests in an apparent dose-response fashion. Figure 1 demonstrates what will be referred to as the "Dose-Response SVT/PVT Effect," in which individuals who pass the SVT/PVT measure on the first attempt constitute the valid group with the highest performance on a neurocognitive measure of memory, the Wechsler Memory Scale –III (WMS-III) Composite Memory Index (Suchy et al., 2012). The scale is in T-score increments, where average is T=50 (standard deviation = 10). All patients had independently been diagnosed with multiple sclerosis (MS) and so the group T-score of ~42 by the MS "valid" group still reflects reduced cognitive ability in the MS group, a finding that would be expected. The graph may be viewed from left-to-right as greater degree of SVT/PVT failure actually influences cognitive performance. Even those MS patients who pass the SVT/PVT on the first round of administration have reduced memory, but increasing levels of SVT/PVT failure relates to reduced memory performance reflecting the Dose-Response SVT/PVT effect.

As shown in Figure 1, from Suchy et al. (2012), comparing groups with different external SVT/PVT scores that are either above (i.e. a "pass") or below (i.e., a "fail") a designated cut-score relates directly to performance levels on separately administered cognitive measures (WMS-III in the study by Suchy and colleagues) where the degree of SVT/PVT failure also relates to the level of memory performance. This well-designed study by Suchy and colleagues is unique in the SVT/PVT research literature because it studied a large sample (N = 507) of individuals who had been independently diagnosed with MS. So the MS diagnosis was not in question and the neuropsychological test findings were being used for follow-up and treatment recommendations. The clinical circumstances of the study combined with the absence of litigation minimized the presence of traditional factors associated with secondary gain. Nonetheless, 11 percent of these clinically referred MS patients failed the SVT/PVT measure administered. When the group who failed the SVT/PVT were confronted and given a second chance, several improved their SVT/PVT performance and scored above the cut score. This

resulted in an overall improvement in their memory whereas those who "failed" the SVT/PVT and did not improve when given a second chance exhibited the lowest memory scores.

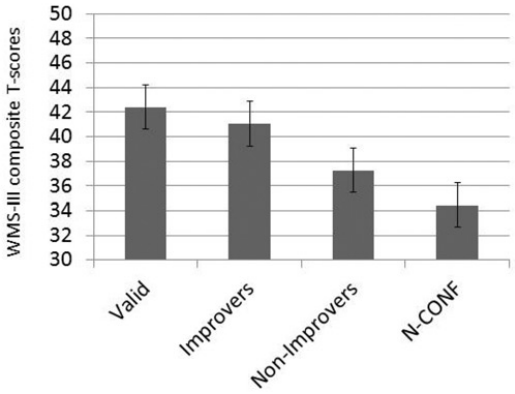

FIGURE 1 Mean composite WMS-III T-score for individuals in four groups: Valid = patients who produced valid Victoria Symptom Validity Test (VSVT) performance during initial administration Improvers = patients who initially produced non-valid VSVT performance but the repeat administration after confrontation yielded valid results; Non-improvers = patients who produced non-valid VSVT performance initially and again after confrontation; N-CONF = patients who initially produced non-valid VSVT performance but were not confronted.
SOURCE: Suchy et al., 2012, p. 1305.

In a different study, Stevens and colleagues (2014) examined the influence of passing or failing a SVT/PVT measure on a clinical sample of individuals with a diagnosis of schizophrenia, most of whom were in residential placement. All of the participants volunteered for the study, and their diagnosis of schizophrenia was already established and independent of any of the SVT/PVT neuropsychological testing performed for this investigation, thereby minimizing any traditional secondary gain issues. Twenty-six percent of the patients scored below the SVT/PVT cut-point, forming the SVT/PVT "fail" group. Figure 2 demonstrates that on all but one measure those patients with schizophrenia who "fail" (depicted in red on the graph) perform less well than otherwise matched patients with schizophrenia who "pass" the SVT/PVT measure. What is striking about this plot is that the pattern of impaired neuropsychological performance is basically mirrored in those who "fail" the SVT/PVT, only the magnitude of the deficit that differs. Also note that there are tests for which there are no differences, indicating that whether or not a SVT/PVT was passed was irrelevant to performance outcome. In schizophrenia there are presumed aberrant neural systems that likely affect

motivation and task engagement along with memory and executive functioning, which are the areas in which distinct deficits were observed even in the schizophrenia group who

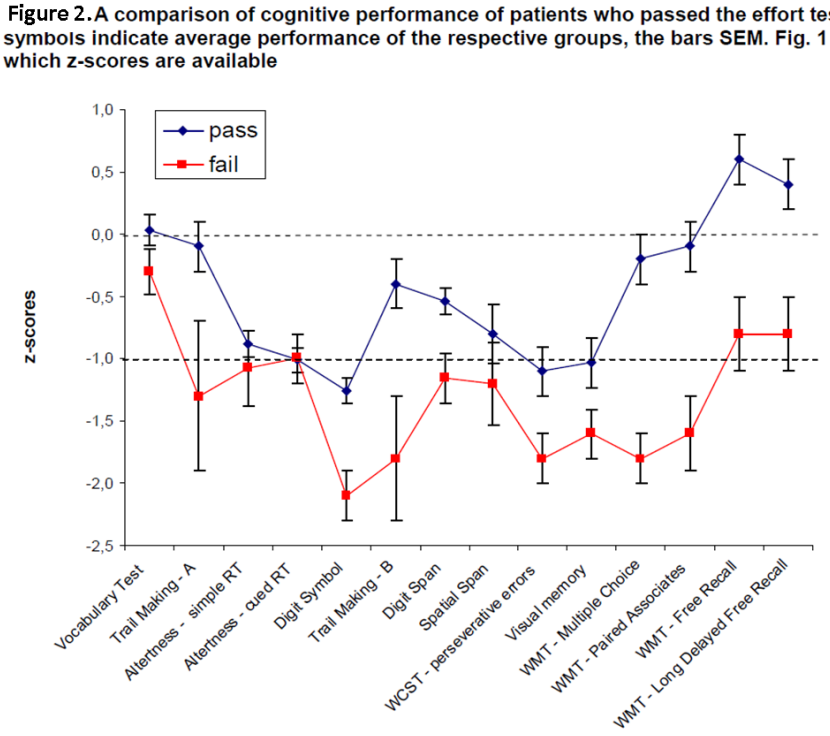

FIGURE 2 A comparison of cognitive performance of patients who passed the effort test with those who failed. The symbols indicate average performance of the respective groups, the bars SEM.
SOURCE: Stevens et al., 2014.

"passed" the SVT/PVT measures. This raises the question of whether SVT/PVT "failure" in schizophrenia is not a real failure but rather a manifestation of the SVT/PVT tapping a dimension of a core deficit that affects cognition and behavior.

There is a substantial list of SVT/PVT studies showing similar dose-response relations and what appear to be less than optimal performance in those who do not "pass" the SVT/PVT measure (Barr, 2013; Bianchini et al., 2001; Bush et al., 2005; Deright and Carone, 2013; Heilbronner et al., 2009; Sollman and Berry, 2011; Vanderploeg and Curtiss, 2001). This literature emphasizes the need for psychologists and neuropsychologists to directly address issues of test validity for any type of assessment performed.

However, there are no specific guidelines for which SVT/PVT should be administered and under what conditions or how to interpret the results. In addition, although the "pass/fail" or "valid/invalid" dichotomy appears simple and straightforward at the individual level, as will be

shown, there are major limitations, complexities, and unknowns when it comes to clinical decision making. In the two examples thus far, SVT/PVT failure rates ranged from 11 percent to 26 percent but in patients with schizophrenia and MS there are likely genuine neuropathological explanations for why there are such high "failure" rates. Indeed, many probably are not truly invalid test performances because the SVT/PVT measure is actually tapping some aspect of a deficit. Another recent study by Hampson and colleagues (2014) investigated inpatients with traumatic brain injury (TBI) along with outpatient community-based TBI patients and a group of patients with chronic epilepsy. This study was undertaken in the United Kingdom and therefore some of the healthcare, insurance, and secondary gain issues that are often written about in SVT/PVT studies were minimized in this investigation. Seven different PVT/SVT measures were administered. One of the most commonly used SVT/PVT measures had a failure rate of approximately 30 percent. In the epilepsy group, a most interesting observation was that one measure was passed by all 16 patients with intractable epilepsy yet 37.5 percent of those same individuals failed another SVT measure. As will be addressed later in this review, the comparability of different SVT/PVT measures is unknown along with how many should be administered and at what point during an assessment the SVTs/PVTs should be administered. In the study by Hampson and colleagues, a patient's performance could be deemed "valid" or "invalid" depending on which SVT/PVT measure the patient received.

All three studies (Suchy et al., Stevens et al. and Hampson et al.) evaluated patients with independently identified disorders that affect cognition, emotion, and behavior. Individuals with disorders like MS, chronic epilepsy, and TBI may qualify for SSA disability depending on the merits of the medical condition at the time of application, but how should a SSA examiner view a "failed" SVT/PVT finding in cases where a bona fide disorder is present? If the SVT/PVT is tapping an element of the disability, then the high failure rates reported (11 to more than 30 percent) represent false-positives. The schizophrenia sample is particularly troubling with respect to drawing erroneous conclusions because the high SVT/PVT failure rates are likely inherent to the disorder. As Harvey and colleagues (2012) have shown there is legitimate functional impairment in people with schizophrenia, including deficits in drive and motivation, and they often perform poorly on measures that assess mental effort and executive function. Harvey and colleagues (2012), p. 1) write: "Research evidence suggests that disability applicants with a valid diagnosis of schizophrenia have significant impairment across multiple dimensions

of functioning, and will typically remain impaired for the duration of normal working ages or until new interventions are developed." SSA grants disability status to a substantial number of individuals with schizophrenia. As stated by Harvey a "valid diagnosis" is key to this determination, but if individuals with schizophrenia exhibit poor task engagement and mental effort as core symptoms/problems of the disorder they have a higher likelihood of performing poorly on SVT/PVT measures. In schizophrenia a failed SVT/PVT score could be misinterpreted as indicating invalidity when it is merely reflecting a core symptom or problem of the illness.

The study by Hampson and colleagues (2014) also highlights the implausibility of using any strict SVT/PVT criteria for SSA disability determination. None of the patients was in litigation. The epilepsy group all had independently diagnosed chronic epilepsy characterized as "intractable epilepsy." The community-based brain injury group resided within a "residential community services" program run by a registered charity, and the TBI patients were either post-acute rehabilitation patients or more chronic community-based patients all of whom had sustained a brain injury sufficient to meet criteria for the UK National Health Services policies for such treatment. By medical history standards, especially the residential TBI patients and the individuals with epilepsy likely would meet SSA disability standards, but what would SSA examiners do with the information about "failed" SVT/PVT performance of 30 to 60 percent in these patients with neurological deficits?

There is another major incongruity over what "passed" SVT/PVT findings mean when evaluating cognitive conditions associated with epilepsy and potentially other conditions (see Bigler, 2014). For example psychogenic non-epileptic seizures (PNES) is thought to be a functional disorder, although it may co-exist with bona fide epilepsy. On neuropsychological tests, PNES patients as a group often perform worse on cognitive measures than do controls without epilepsy. Likewise, PNES patients complain of cognitive impairment and although they may have a functional disorder, what does it mean when PNES patients "pass" all SVT/PVT measures administered (Dodrill, 2008; Drane et al., 2006; Williamson et al., 2012)? Does a PNES patient who passes SVT/PVT measures validate and substantiate a cognitive impairment by virtue of passing so-called validity measures?

The Suchy, Stevens, and Hampson studies highlight the complexities of the issues surrounding the interpretation of SVTs/PVTs. All three studies minimized overt secondary gain issues and were conducted in clinical samples where the psychological/neuropsychological

testing was not done for disability determination. Nevertheless, high SVT/PVT "failure" rates were reported. High SVT/PVT failure rates occur in a number of clinical populations, as will be discussed below.

Unfortunately, anyone performing a disability assessment is confronted with the role that deception plays in human behavior (Ford, 2006; Poole, 2010). Only genuine disorders that are truly disabling should qualify for such a designation. Accordingly, it is important to accurately identify the charlatan but not at the expense of misclassification of those with genuine disorders and disability(Wasyliw and Cavanaugh, 1989). As will be discussed, the current methods of psychological and neuropsychological assessment are the best measures of the day for evaluating presence of cognitive, behavioral, and/or emotional disorders. Some application of SVT/PVT findings do appear to be relevant for disability determination, but in terms of current practices there are many limitations, exceptions, and problems, as will be identified in this review.

The following section contains a general statement about the use of SVT and PVT measures in standard disability assessment. To more fully understand and address the issues raised in the general statement, a historical perspective on the development of validity measures in clinical settings is then provided, followed by a discussion of various SVT/PVT issues, critiques, and unresolved questions. The paper concludes with individual responses to each of the questions posed in the Institute of Medicine (IoM) request. I have written two general reviews on SVT/PVT measures (see Bigler, 2012, 2014) and for efficiency, the reader is referred to those publications as background. Since this IOM committee has asked for my opinions, some of the following text is written in the first person.

GENERAL STATEMENT

As of 2014, the American Academy of Clinical Neuropsychology (AACN) and the National Academy of Neuropsychology (NAN) have endorsed the concepts of external SVT/PVT measures as an important aspect of any neuropsychological examination (Bush et al., 2005; Heilbronner et al., 2009). In other words, to meet current accepted practice standards, psychologists performing evaluations that assess cognitive and emotional functioning for diagnostic purposes need to comment on the validity of test findings and need to use empirical methods. Merely offering a clinical statement or personal judgment based solely on observation

that test results appear to be a valid representation of the examinee's ability does not meet recommended practice standards.

A passed SVT/PVT represents the best marker to date for commenting on the overall validity of a psychological/neuropsychological assessment. Similarly, for SVT/PVT measures that use a forced choice (FC) format, the binomial theorem and the probability of a chance occurrence can be applied to any score. SVT/PVT scores below chance become sine quo non indicators of invalid performance. If there are only two choices purely random responding centers at 50/50 chance occurrence with anything deviating from that in either direction indicating non-chance. In substantially below chance SVT/PVT scores, the only plausible occurrence of such a response set would be selection of the incorrect response despite knowing the correct answer. In such a circumstance, most likely the individual knows the correct answer and falsifies their performance by selecting the incorrect.

As will be explained in this review, in a typical clinical sample the majority of individuals pass SVT/PVT measures allowing the examiner to conclude that valid test performance was present if there are no other contraindications. Usually only a small minority will perform at or below chance levels (Bianchini et al., 2003) and therefore unequivocally support a conclusion of test invalidity. For these two circumstances straightforward commentary about test validity can be made, where SSA examiners could generally assume valid test findings when the SVT/PVT score is at or above a designated cut-score and invalid when below chance SVT/PVT performance is documented.

However, as reviewed by Bigler (2012, 2014) a substantial minority of individuals, in some clinical settings 30 percent or more, technically "fail" SVT/PVT tasks by performing below the established cut-point, although their scores are above chance, and some are near misses, just one or a few points below the established cut-score. For failures below the cut-point but substantially above chance, there is insufficient outcome research to know on an individual basis how to objectively interpret such findings as valid or invalid. At this point, most SVT/PVT measures do not permit flexibility in terms of cut-scores; therefore, if 45/50 is the established score for a "pass" then a score of 44/50 constitutes a "fail."

There are no SVT/PVT studies that have been funded by government or foundation grants that independently assess the effectiveness of SVT/PVT measures where a specific type of neurological and/or neuropsychiatric disorder is systematically examined. There are no large

scale studies that have examined SVT/PVT findings using independent lesion markers, such as those identified via neuroimaging methods, to examine the influence of lesion type, burden, or location on SVT/PVT findings. There are no large-scale systematic studies that have examined whether there should be variable cut-points depending on the clinical characteristics and history of the individual being tested, including any medications the person is taking. The issues surrounding Type 1 versus Type 2 statistical errors as they relate to clinical decision making and disability determination have not been adequately addressed in the SVT/PVT literature. The absence of more definitive understanding of classification error in SVT/PVT measures limits their universal application, especially in disability determination. As outlined by Young (2014), there are more than 40 SVT/PVT measures and techniques in current use, but no head-to-head comprehensive study has been conducted comparing the effectiveness of different SVT/PVT measures and their interpretation across diseases and disorders.

One core problem and interpretive limitation with SVT/PVT findings is demonstrated in an outcome study by Locke and colleagues (Locke et al., 2008). None of the participants were in litigation and some had already received disability status, so there was no external motivation, per se, for compensation seeking. All participants were being evaluated for treatment planning in an outpatient rehabilitation setting, yet about 20 percent failed the SVT/PVT used by these clinicians. Although at a group level those who failed the SVT/PVT measure generally performed worse than those who passed (a robust finding consistent with the dose-response SVT/PVT performance effect previously described), group differences between those who "passed" and those who "failed" were not a universal finding. On two measures of executive functioning (Halstead Category Test and Wisconsin Card Sorting Test), the individuals who failed the SVT/PVT measure did not differ from those who passed on almost all demographic variables, including age, education, psychiatric history, or years since injury. The only difference was higher levels of disability in the group that failed the SVT/PVT measure. Since disability had already been independently established, does this finding mean that because there were more individuals previously assessed to be disabled among those who failed SVT/PVT measures that the SVT/PVT was tapping their level of disability? Or does it mean that they may have feigned their deficit in the first place? The answer to these questions is simply unknown. Whether a group passed or failed the SVT/PVT made no difference with regard to executive measure test scores. So what grounds are there to conclude that the executive test results were not valid in the

group that “failed” the SVT/PVT measure but were valid in the group that “passed”, since there was no direct measure of test validity during administration of the executive function measure? The problem with external SVT/PVT measures is that, by definition, they are never administered during the actual test for which validity is being assessed.

A typical neuropsychological evaluation includes multiple tests and subtests taking several hours to administer, and some evaluations may involve an entire day or more than a day. An external SVT/PVT is supposed to act as an omnibus measure indicating the validity of all tests administered, but the actual validity of a single test administered may never be directly assessed, unless it includes a proven internal SVT/PVT metric that can be examined. So determining whether a complete battery of test results is valid or not is merely an inferential process, but should it be solely based on a SVT/PVT measure?

Over the last 25 years, there has been a proliferation of external SVT/PVT measures instead of development of new psychological and neuropsychological assessment measures with internal validity checks. Before reviewing some of the historical roots of SVT/PVT development, a general conclusion will help to frame the discussion.

Preliminary Conclusion: SVT/PVT interpretation is a qualitative process. There are certain situations, namely “passes” and below chance failures, in which SVT/PVT findings lead to definitive conclusions about validity of psychological or neuropsychological test results. Current professional standards of practice require the validity issues to be empirically addressed as best as possible but guidelines are lacking. The inherent problem with current psychological and neuropsychological tests is that individually administered tasks lack internal markers of validity, although some permit post-hoc analysis of some embedded test characteristics. External SVT/PVT measures provide some metrics to comment on validity, but there are unacceptably high rates of failure in some groups of individuals with legitimate diseases or disorders that eliminates any universal application of a single SVT/PVT method or metric, especially with strict cut-point rules. The real message of SVT/PVT research is that there are fundamental flaws in the current methods of psychological and neuropsychological assessment that need to be corrected, emphasizing the need for new testing methods with embedded validity measures within each test or test domain. This would permit abandoning the omnibus external SVT/PVT approach as the inherent limitations of such measures are likely insurmountable as will be shown by this review.

HISTORICAL ROOTS OF SVT/PVT DEVELOPMENT

Sage Advice from the 1960s

Beginning in the twentieth century, psychological assessment techniques and the development of standardized ability testing have played critical roles in education, health care, business and industry, and government policies. In response to the increased use of ability testing to classify and potentially segregate individuals based on their testing profile, almost half a century ago Goslin (1968) wrote an extensive critique of standardized ability tests and their use and interpretation. In my opinion, the review by Goslin represents a "Gold Standard" critique for the field that is applicable to the discussion of SVT/PVT use. The abstract of Goslin's review captures the essence of the major points and is reproduced as follows:

> At the outset a distinction was made between criticisms directed at the validity of tests and criticisms not affected by the validity of the tests. It was noted further that all criticisms of tests must take into consideration the type of test and the use to which the test is put. Criticisms of the validity of tests involved the following issues: (i) tests may be unfair to certain groups and individuals, including the extremely gifted, the culturally disadvantaged, and those who lack experience in taking tests; (ii) tests are not perfect predictors of subsequent performance; (iii) tests may be used in overly rigid ways; (iv) tests may not measure inherent qualities of individuals; and (v) tests may contribute to their own predictive validity by serving as self-fulfilling prophecies. Criticisms that are more or less independent of test validity included the effects of tests on (i) thinking patterns of those tested frequently; (ii) school curricula; (iii) self-image, motivation, and aspirations; (iv) groups using tests as a criterion for selection or allocation, or both; and (v) privacy. Several concluding remarks are in order: (1) This paper has focused almost entirely on criticisms of tests. However, the positive value of standardized tests should not be ignored. Here we must keep in mind what possible alternative measures would be used if standardized tests were abandoned. (2) We must begin thinking about tests in a much broader perspective—one that includes consideration of the social effects of tests as well as their validity and reliability. (3) Finally, an effort should be made to develop rational and systematic

> policies on the use of tests with the culturally disadvantaged, the dissemination of test results, and the problem of invasion of privacy. Such policies can be formulated only if we are willing to take a long hard look at the role we want testing to play in the society. Standardized tests currently are a cornerstone in the edifice of stratification in American society. It is up to the social scientist to conduct research that will enable policy makers in education, business and industry, and government to determine in a consistent and rational way the ultimate shape of this edifice. Goslin (1968)

The criticisms of psychologically based testing as outlined by Goslin, in part, frame the review that follows on the use of SVT/PVT measures. A year after Goslin's critique came the publications by Nussbaum and colleagues (Nussbaum et al., 1969a; Nussbaum et al., 1969b) implicating the important role that psychological assessment techniques could play in Social Security disability determination. These publications occurred long before major work began on SVT/PVT measures and their clinical application. On the one hand, as indicated by Nussbaum and colleagues, psychological testing should provide a wealth of important information for the disability examiner. On the other hand, as pointed out by Goslin, fundamental assessment issues both psychometrically and clinically need to be scientifically addressed.

SVT/PVT Foundations

External SVT/PVT Test Development

As summarized in Bigler (2014), André Rey, the author of the Rey Auditory Verbal Learning test (RAVL), introduced the RAVL in the 1940s, and it soon became one of the most widely used measures of verbal list learning (Lezak et al., 2012). Rey quickly realized the need to validate memory performance understanding that the examinee's response was dependent on cooperation and effort. The mid-twentieth century was the zenith for discussions that centered on neuroses and non-conscious motivations. Rey was well aware of so-called functional disorders that could mimic neurological conditions. The quest in this era was to use a psychological assessment technique that could differentiate "organic" pathology from "functional disorder," implying a non-organic basis to the condition. Rey is credited with introducing the "easy task" SVT/PVT approach wherein the cognitive demands of the task were trivial and easily passed by

most individuals, even those with significant neurological and neuropsychiatric disorder. An example is shown in Figure 3. This type of format was the origin of the Rey-15 Item Test (Rey-15) that was introduced in the mid-twentieth century and discussed in his 1958 book (Rey, 1958). The instruction format given to the examinee follows something like this (not the actual instructions, nor actual test): “In this next test I am going to briefly show you a card (see Figure 3) with symbols, letters and numbers. I want you to study what is on the card and then from memory reproduce what you can. Here is the card but you can only briefly view it.”

Rey assumed that if basic test engagement could be demonstrated on an easy formatted task like this, it would generalize to task engagement for other aspects of test performance. Note the simplicity of the task but it has face validity as a memory test and it has 15 items that have to be retained. For the individual who is not test savvy, this has every appearance of being a legitimate memory test, but as shown by Rey (and numerous others who have studied the Rey-15 item or similar tasks, see Reznek, 2005; Whitney et al., 2008), typical developed controls universally have no trouble achieving near to perfect scores on this task as do the majority of patients with a significant neurological disorder. With a cut score set at something like 9 of 15 items retained, poor performance defined as a score below the cut-point on this measure suggested invalid test performance on other tasks administered, supporting the idea that an external measure could be used to comment on the validity of other tests administered in the same session (Bernard and Fowler, 1990; Goldberg and Miller, 1986).

+ - X

X - +

A B C

a b c

1 2 3

FIGURE 3 A prototype example of an “easy” SVT/PVT format memory task. Although 15 items are presented, the actual content that needs to be retained is minimal.

Importantly, because measures like the Rey-15 appear to be actual memory tests and because they are intermingled with other tests during an evaluation, the examinees assume they are in fact performing a task that measures a particular cognitive ability. They are not informed about the ease and trivial nature of the memory task or that its intent is to measure validity of

performance. Rey used a cut-point approach and even though the task was easy and the majority scored close to perfect or without error, a liberal cut-point (e.g., 9 items correct out of 15, see Hays et al., 1993) gave the benefit of doubt to the examinee. Since the Rey-15 does not contain words and given the trivial nature of the task, its presentation did not interfere with word retention from the RAVL. Giving the Rey-15 and having the patient "pass" it, provided greater confidence to Rey and others who used his test that an individual's memory deficits were "organic" if the person performed poorly on the RAVL and the clinical history merited such a distinction. Note that the Rey-15 is an external measure but traditionally was administered in temporal proximity to the RAVL, which permitted direct comment on the RAVL findings. Thus, the external SVT/PVT was established. It is also important to note that in this era there were but a handful of basic ability tests that could be administered. The Rey-15 was introduced prior to the development of the lengthy battery approach to neuropsychology that emerged in the 1960s.

Although the Rey-15 Item was an external SVT/PVT, it was used by Rey to comment on just a limited number of cognitive measures. In contrast, participants in the previously mentioned study by Locke and colleagues (Locke et al., 2008) underwent an entire day of assessment, yet validity was based on the performance of a single external SVT/PVT. In the Locke study, the validity of the findings from approximately 36 index or summary scores was inferred based on a single pass-fail external SVT/PVT measure.

The ubiquitous role that memory plays in all aspects of cognitive assessment and that certain memory assessment formats are readily adapted for SVT/PVT measures, especially the forced choice format set the stage for development of other validity measures over past three decades (Bernard, 1990; Bernard and Fowler, 1990; McGrath et al., 2010). Not all cognitive SVT/PVT measures use a memory format, but most do, beginning with the Rey-15 Item. Since there is not a cognitive assessment task that does not tap some aspect of working memory, it is not surprising that this SVT/PVT memory format has been an effective approach.

The basic SVT/PVT premise for cognitive measures is as follows: Instructions either oral or visual have to be attended to (held in working memory) and sufficiently retained long enough to be responded to, even if the task is a simple motor response to some stimulus. With increased task complexity, more recruitment of various neural networks that require memory and processing speed are required, so to minimize this, SVT/PVT measures are kept simple and basic. Memory for cued recall, especially after either a short-delay or immediately after stimulus

presentation, is superior to free recall, a well-established and replicated fact since the beginnings of cognitive psychology research from late nineteenth and early twentieth century (Jung, 1967). Combining the "easy task" approach with some aspect of a cued recall task, where one item (target stimulus) has been viewed (or processed in some fashion) paired with another item that has not (foil stimulus), creates a forced choice paradigm. Furthermore, if the numbers of items is kept to 50 or less, forced choice recognition memory with no delay is relatively error free in typically developed children and adults, a known fact from cognitive psychology that has endured more than a century of investigation (Glanzer et al., 1993).

In a forced choice paradigm, if the task is designed to be easy to perform then the majority of those who take the test perform close to or with 100 percent accuracy. All commercially available forced choice SVT/PVT tasks have this in common and all report normative findings in control participants to readily perform with few to no errors. This permits the setting of high criterion levels for cut-scores typically in the 85 to 90 percent accuracy range. If easily passing a forced choice task is the norm, then failure may signify deficient effort or poor task engagement making the results unreliable.

The potential dilemma is that despite the ease of a task, some individuals may fail to reach the pass-fail cut score because of factors associated with their disability. Although there have been statements made in the SVT/PVT literature that these tasks require "no effort," that is simply incorrect (see Bigler, 2014). All commercially available SVT/PVT measures have included a "mixed" neurological group typically composed of patients independently assessed to have suffered a stroke, to have some form of acquired brain injury, or to have some medical condition known to compromise cognitive ability. These neurological groups are heterogeneous with none of the clinically available SVT/PVT measures systematically examining neurological conditions by category and most do not provide any method for moving the cut-point based on the condition being assessed. Although all commercially available SVT/PVT measures discuss the potential for false positive classification in the midst of SVT/PVT failure, without knowing how SVT/PVT performance may vary across disorders and how adjustments may be necessary, the clinician is left with few guidelines on how to identify or handle false-positive findings. Some SVT/PVT measures have recognized this problem and attempt to correct for it by a "profile" analysis, but it should be apparent that if one has to apply a profile analysis to a

SVT/PVT measure, it is doing something more than just tapping a valid versus invalid dimension.

Internal or Embedded SVT/PVT Development

Beginning in the early 1970s, Elizabeth Warrington began the systematic study of recognition memory of faces and words associated with normal aging and dementia along with the influence of lateralized cerebral pathology (Warrington, 1974; Warrington and Taylor, 1973). The initial premise was that deficits in forced choice word recognition would be associated with left hemisphere pathology and impairment in forced choice face retention would be a marker for right hemisphere damage (however, see Morris et al., 1995). In 1984, Warrington released the Recognition Memory Test (RMT) for clinical use. The RMT used a recognition format to assess retention of 50 unfamiliar faces seen or 50 common words. The interesting observation made by Warrington was how relatively easy the task was for individuals in the control group and even for some neurological patients who did not have temporal lobe pathology or pathology that would influence perceptual or language processing. Although it may seem a rather daunting feat to retain 50 faces or words viewed for only 3 seconds, in fact the intact brain in the individual with typical development has no difficulty with this kind of a task, as the brain is readily primed to discriminate and retain faces and words (Standing, 1973).

Soon after the Warrington RMT was released, discussion in the clinical literature began about potential validity indicators that could be derived from it (see Nelson et al., 2003). On the retention trial, regardless of whether immediate or delayed, the target stimulus was either a previously seen face or word, paired side-by-side with a word or face that had not been seen previously—the foil stimulus—where the individual had to choose which item they believed had been seen previously. The simplicity of this forced choice approach from the perspective of developing SVT/PVT indices is that if one were to "guess" there is a 50/50 chance of being correct if truly guessing (the target and foil stimuli are counter balanced as either being on the left or right on the stimulus presentation card). In a bona fide amnestic patient with profound short-term memory impairment, random responding would generate ~25 correct responses. However, in the presence of feigning, scores of less than 25 would suggest non-random responding, in which the incorrect answers were intentionally selected.

As originally developed, there was no intention for the Warrington RMT to be a SVT/PVT measure, but its format was readily suited to function as both a clinical and SVT/PVT measure; in other words, it has its own built-in, internal or embedded SVT/PVT. Since it is a 2-item forced choice neuropsychological measure of memory, binomial probability can be applied.

In bona fide memory disorders with documented temporal lobe pathology, such as in Alzheimer's disease, patients perform poorly on the Warrington RMT in proportion to the degree of temporal lobe pathology (Cahn et al., 1998). Profound bilateral temporal lobe pathology may produce a dense amnesia where RMT performance occurs at chance levels. Such a case was shown by McCarthy, Kopelman, and Warrington (2005). Figure 4 depicts the obvious bilateral temporal lobe pathology.

For this case reported by McCarthy and colleagues during one testing session with the Warrington RMT the score was 25/50 for words and 31/50 for faces. Technically, from a SVT/PVT perspective, this represents a "failure."

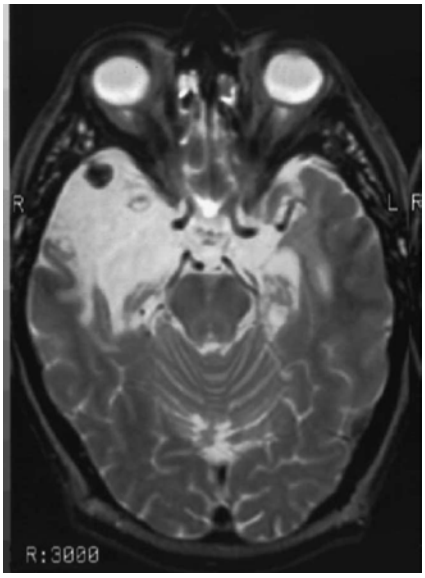

FIGURE 4 MRI Scan of RFR at the level of the temporal lobes. MRI scan showing extensive bilateral damage to the medial temporal lobes together with involvement of the anterior and lateral portion of the right temporal cortex (left hemisphere is shown on the right of this figure).
SOURCE: McCarthy, Kopelman, and Warrington (2005).

Since Kim and colleagues (2010) established a cut-score on the Warrington RMT of $\leq$ 42 as indicative of what they termed "non-credible" performance, such a classification becomes problematic in any patient group with suspect or documented mesial temporal lobe or

hippocampal pathology and bona fide memory disorder, such as the case presented in Figure 4. Other research on the Warrington RMT has shown how an examinee's level of intellectual ability or type and extent of neurological deficit present may affect performance (Dean et al., 2008; Dean et al., 2009; Nelson et al., 2003; Smith et al., 2014). This demonstrates the problem with so-called "hard rules" with respect to cut-scores and their interpretations if a measure is to be used as a SVT/PVT. As with the case presented by McCarthy and colleagues (2005), given the positive MRI findings and history of herpes simplex encephalitis, this level of performance is exactly what would be expected on a forced choice two-item recognition task in a patient with massive bilateral mesial temporal lobe damage. In cases like this, the neuroimaging findings and medical history trump anything that could be concluded about the "validity" component of the Warrington RMT other than that it was validly assessing the cognitive deficit (in other words a false-positive marker of invalidity). More flexible approaches to SVT/PVT interpretation measures like the Warrington RMT may be of great value in making interpretive statements concerning test validity and clinical correlations because it functions as a dual measure – both clinical assessment of memory and an embedded approach to validity. Unfortunately, there are few measures like the Warrington RMT that are being used and even fewer studies that have even attempted to examine neuroimaging findings related to SVT/PVT findings with cut-score adjustments (see review by Bigler, 2014).

The other problem with embedded measures is that their use as SVTs/PVTs is based on post-hoc analyses and comparisons, since they were not originally devised to function as such. Accordingly, the ability of embedded measures also to detect failed SVT/PVT performance for externally administered tests is often less than satisfactory (Meyers et al., 2011; Miele et al., 2012). Accordingly, there are no universal recommendations for simply using psychological and neuropsychological measures that have embedded SVT/PVT metrics.

Specific Sensory Deficit, Disability Determination, and SVT/PVT Measures

All aspects of SVT/PVT tasks require that the examinee have minimal sensory and basic processing along with motor/language ability to perform the elementary requirements of test taking. Nonetheless, it is not uncommon for a patient undergoing some aspect of assessment to have a hearing, sensory, or motor deficit for which they are seeking compensation. In some respects, the forced choice SVT/PVT is ideally suited to assess these kinds of specific deficits, in

part because the sensory SVT/PVT task is assessing both the potential deficit and simultaneously the validity of the measure. In this circumstance the SVT/PVT is functioning diagnostically (truly assessing whether a deficit is present) as well as providing information about whether the response is valid. Built-in validity and diagnostic capability should be the model for all future ability and symptom-based measures.

In these specific sensory and perceptual circumstances, the forced choice method may be applied to signal detection techniques that identify malingered or hysterical sensory loss involving any sensory modality. As demonstrated by Pankratz, Fausti, and Peed (1975) these techniques involve application of a sensory stimulus that the examinee claims an inability to perceive. The examinee is then instructed to "guess" about some characteristic of the stimulus that they claim they cannot recognize (e.g., whether the stimulus is present or not, how long the stimulus is on, stimulus intensity, etc.) which if they truly could not perceive would result in random guessing and chance performance. However, regardless of whether performance is above or below chance levels, either would indicate some ability to recognize the sensory stimulus. The limitation with techniques like this is the inability to determine whether there is conscious intent to malinger or whether the response-set is driven by hysterical, somatic/conversion symptomology (Pankratz, 1979).

Additional Relevant Background and SVT/PVT History

Given the foundational studies reviewed above (see also Young, 2014), from the late 1980s on, numerous SVT/PVT tests and approaches have been published and applied clinically. The simplicity of the tasks and/or the unlikely combination of symptoms or performance levels seem to make clinical sense as indices of validity or non-validity, but do current SVT/PVT measures provide the clinician with the kind of universal validity assurance that some proponents of SVT/PVT endorse? In other words, how confident can the clinician be in making SVT/PVT interpretations? Some additional historical perspectives should assist in answering these questions.

The first standardized psychological assessment methods began more than a century ago and primarily were designed to address educational and vocational training needs. Intellectual and ability testing (see Anastasi, 1954) led the way, followed by so-called personality, vocational, and personal interest questionnaires. The Stanford-Binet Intelligence Scales still in

use today (5th Edition from the original Binet) had its origin in the first decade of the twentieth century. The Wechsler scales of intelligence date their origin to the 1930s, with a fifth edition of these intelligence tests currently in development. As currently formulated and normed, neither of these venerable measures ever focused on or included rigorous embedded validity methods. All validity statements that can be made about these intellectual measures come from post-hoc derived methods (Davis et al., 2011; Harrison and Armstrong, 2014; Williams, 2011; Yang et al., 2012). So part of the twenty-first century dilemma about how to assure test validity is rooted in the fact that internal or embedded validity measures within each test were never initially established. Some twenty-first century assessment methods remain basically unchanged from 100 years ago.

As outlined by Anastasi (1954) , the historical validity concerns of the day were focused on the traditional test development issues of internal consistencies, construct validity, test-retest reliabilities, convergent and discriminant construct validity, and the like, assuming that individual test engagement and sufficient effort (motivation?) to perform would occur because of the demand characteristics of the assessment environment. Of course, cautionary statements were made about gaining the examinee's cooperation and the tests being deemed invalid should the examinee either not be paying attention or be exhibiting outwardly non-compliant behavior. The tests were not designed to assess sub-optimal effort or other reasons why test performance may be subpar. The original intellectual assessment measures were all based on individually administered tasks where it was assumed the examinee would naturally perform to her/his best ability being in the presence of an examiner and any noncompliance would be detected by the examiner. In other words, the demand characteristics of the one-on-one test environment were considered sufficient, presumed to have ensured validity of the examinee's test performance such that no internal validity checks were needed. In the beginning of psychological test development, these assumptions were never empirically tested but only assumed to be true. Although in the 1960s, as already reviewed, Rey began to empirically demonstrate some of these limitations, it was not until the 1970s and 1980s when the SVT/PVT movement really began to systematically address issues of test validity with standardized measures that had no internal validity checks (Binder and Pankratz, 1987; Pankratz et al., 1987).

Simultaneous with the development of individually administered cognitive and ability tests during the first decades of the twentieth century, so-called personality assessments,

symptom reporting, and interest inventories began to be developed. These measures were completed by the individual, typically with either a Yes/No or True/False checkmark (Anastasi, 1954; Hase and Goldberg, 1967), and there was no need for the clinician or examiner need to be present or to ask the questions. Since there was little or no examiner involvement in the administration of these tests, test developers assumed from the beginning that some index of the veracity of reported or endorsed symptoms/problems was needed. The oldest and most widely used measure of personality and emotional functioning, the Minnesota Multiphasic Personality Inventory (MMPI, Hathaway and McKinley, 1943), dates its origin to the 1930s, becoming commercially available in the 1940s (Hathaway and McKinley, 1940). The MMPI became the prototype for another SVT approach. In developing the MMPI, hopeful that it would have utility even in assessing individuals with thought disorder and psychosis, Hathaway and McKinley, original authors of the MMPI, knew that multiple indexes of the validity of MMPI item responses were essential (see Hathaway, 1947). Hence, MMPI validity scales were developed. These validity scales use a rather simple premise – not responding (number of blank, skipped, or non-responded items), unusual responding, responding all true or all false, or responding in ways that overly emphasized certain attributes or symptoms would raise suspicion about the accuracy of the response set of the individual completing the test. Not only could a response set be unlikely, using the normative and statistical approach as provided by Hathaway and McKinley, some statistical likelihood of occurrence in comparison to the typical population, including cases with similar disorders could be established (see Hathaway and Briggs, 1957). Furthermore, early on in the clinical use of the MMPI, clinician investigators were attempting to define MMPI metrics of dissimulation (Gough, 1950). At least from the questionnaire standpoint, assessing "symptom validity" became central to most inventories and a critical first statement forming the basis for interpretation of a personality test.

Some of the next developments in SVT/PVT research came with the observation that "failed" MMPI SVT measures, when used to craft groups of patients undergoing neuropsychological testing, would actually underperform, often significantly so, despite being matched on other demographic variables with a comparison group that passed MMPI validity scales (see Jones and Ingram, 2011; Larrabee, 2003a, 2003b; Temple et al., 2003). Such findings reinforced the idea that questionnaire based SVT measures within tests like the MMPI or other personality measures could generalize to actual cognitive assessment, reflecting invalidity on

cognitive test performance as well. In other words, if someone “failed” SVT measures on a personality test, they would be more likely to demonstrate invalid performance on cognitive ability tests.

As already mentioned, the MMPI had traditional internal validity checks but also identifiable profiles that purportedly were capable of detecting dissimulation if examinees were attempting to portray themselves as better or worse than what was true (i.e. “faking good” or “faking bad”— ("claiming symptoms and problems they really did not have, p.264; Graham et al., 1991). Faking bad on the MMPI was originally determined by having examinees with no disorder attempt to “fake” a mental disorder (Peterson et al., 1989) or simulate how someone with an injury would feign an impairment compared to individuals with the disorder (Lees-Haley et al., 1991). Observing these relations in a forensic sample of personal injury claimants, Lees-Haley and colleagues (1991) developed what they titled a “fake bad scale (FBS)” derived from the MMPI. The FBS examines patterns of symptom endorsement that are potentially implausible in their combinations, levels of endorsement, and frequencies of occurrence.

MMPI related research became the cornerstone for valid and invalid symptom reporting, and various methods purported to detect “dissimulation” characterized by symptom overreporting were published (Hyer et al., 1989). The problem, also addressed by Hyer and colleagues, with so-called MMPI dissimulation indexes was the difficulty discerning individuals with bona fide psychiatric or medical disorders from those feigning a disorder. More refinement in MMPI studies with regard to validity issues and how deception could be detected was needed (see Merydith and Wallbrown, 1991). In an attempt to meet this need, research on the MMPI FBS burgeoned (see Greiffenstein et al., 2004), and, as early as 1997, Larrabee (1997), p. 203) stated “somatic malingering should be considered whenever elevations on scales 1 and 3 exceed T80 (referring to a “T” statistic with a mean of 50 and standard deviation of 10), accompanied by a significant elevation on the FBS.” Ultimately the FBS was incorporated into the MMPI-2 (http://www.upress.umn.edu/test-division/mtdda/mmpi-2-symptom-validity-scale-fbs, accessed January 13, 2015). However, the items that make up the FBS also can be categorized by somatic symptoms, sleep disturbance, stress or coping issues, low energy or anhedonia, and denial of deviant attitudes or behaviors (Butcher, 2010; Butcher et al., 2003). Patterns of response to this array of questions embedded within the MMPI-2 showed that certain cut-points differentiated examinees thought to have somatoform disorder or frank malingering but on the other hand so do

some legitimate cases where pain, injury, neurocognitive and neuropsychiatric disorder coincide. For example, Gass and Odland (2014) in a non-litigation neuropsychological setting including Veterans Administration referrals concluded that the FBS possesses “ambiguous meaning because of its structural discordance” (p. 1, see also Gass and Odland, 2012).

Other refinements to validity testing with the MMPI-2, including a response bias scale (RBS), were developed(Gervais et al., 2007) and compared favorably to the FBS but also potentially provided some unique ability to detect invalidity of an MMPI profile (Gervais et al., 2010; Nelson et al., 2007). Further adaptations of the MMPI-2 occurred to reduce the number of FBS items, with retention of the RBS in the shortened version referred to as the MMPI-2-restructured format or MMPI-2-RF (Wygant et al., 2010). Purportedly the RBS like the FBS is sensitive to overreporting and apparent symptom exaggeration (Bolinger et al., 2014; Jones et al., 2012; Tarescavage et al., 2013). Despite the optimism offered by proponents of MMPI SVT measures and their clinical applications, the real challenge comes when attempts are made to use MMPI data to separate in individual patients true somatoform disorder from somatic malingering and bona-fide medical conditions with significant somatic complaints (Sellbom et al., 2012). This is not a new problem; indeed, this has been an inherent challenge associated with using the MMPI from the beginning (Leavitt, 1985).

MMPI validity measures, or similar indices from other personality tests, may provide meaningful information but require substantial clinical correlation and integration of medical and psychosocial history for interpretation. Although group data from forensic-based studies show significant levels of non-credible findings when MMPI invalidity measures are high, it is much more challenging to make such conclusions at the level of the individual patient. The complexities of what constitutes a medical history and clinical presentation, what may be pre-existing features of a trait that may pre-dispose one to a particular condition or response set to a questionnaire, and how a genuine disorder becomes expressed in an individual would be difficult to capture by a specific score or profile on the MMPI that could universally be interpreted as representing invalidity.

Forensic Driven SVT/PVT Research

Starting in the 1980s neuropsychological assessments became pivotal evaluations in personal injury litigation (Bigler and Brooks, 2009) and forensic issues began to drive SVT/PVT research. Psychological testimony in criminal litigation had a long-standing history but tended to focus mostly on personality assessment measures (Wasyliw et al., 1988). In criminal litigation, cognitive testing began to be used as mitigating evidence for a diminished capacity defense, especially in death penalty cases (Simpson, 2012). In the later part of the twentieth century, challenges as to whom could be put to death began to be raised on the grounds of intellectual and cognitive disability, with a 2002 Supreme Court ruling (*Atkins v. Virginia*) on the level of intellectual ability that must be present before the death penalty could be carried out. This in turn directly implied that validity of assessed cognitive impairments had to be substantiated, furthering the importance for a role that SVT/PVT measures could play in such evaluations (Chafetz and Biondolillo, 2012; Lewis et al., 2004). Accordingly, at both the personal injury and the criminal litigation fronts, SVT/PVT methods became centerpiece in an attempt to improve assessment accuracy.

However, in the adversarial legal world, SVT/PVT findings and research has polarized implications. Although both sides in a forensic case want information about test scores and their validity, how that information is used in a legal setting may be very different from how an examiner may be reporting or interpreting a SVT/PVT finding. How the referral process has influenced publications in SVT/PVT research has not been systematically examined but is likely influential because the legal system controls how referrals are made and remunerated. There are agendas on both sides. For the SSA, some of these legal arguments are germane because they could be relevant to how SVT/PVT findings could be used in disability claims.

There is no better method to discredit a plaintiff's claim in a personal injury case or a defendant's statement of innocence in a criminal case than to use the malingering designation. A separate section in this review addresses the malingering issue, where likely certain SVT/PVT results are, in fact, definitive for malingering. However, most of the time a SVT/PVT score below the cut-score but above chance performance cannot be used by itself to classify malingering. The issues specific to making a diagnosis of malingering are complex and require multiple data points and inferences (see Bass and Halligan, 2014), with the SVT/PVT finding representing only one data point. The simple "pass versus fail" SVT/PVT dichotomy makes it

appear much more straightforward than what it may be in reality(Silver, 2012), but in the courtroom, questions can be asked in ways that appear to inflate what a SVT/PVT finding may imply.

This can be readily demonstrated by the following line of legal questioning. In this scenario the attorney wants to extract a conclusion that a SVT/PVT finding provides the legal evidence to establish a malingering diagnosis, even though the only finding may be that a SVT/PVT measure has been performed below a cut-point:

ATTORNEY: Dr. XX –"On the _ _ _ performance ***Validity*** test you administered, the examinee ***FAILED*** that measure, correct?

Psychologist: Yes

ATTORNEY: Invalid performance on a test may be a sign of Malingering, is that not correct?

Psychologist: Yes

Attorney: Consider this a hypothetical question and please answer accordingly. If someone is malingering and generates invalid test results, the test results cannot be believed, correct?

Psychologist: Yes

As can be seen by the dialogue presented above, the psychologist never has to really answer the question whether the examinee/patient is malingering, but the line of questioning clearly raises the specter of non-believability. By the end of the twentieth century and beginning of the twenty-first, a psychologist making a statement in the courtroom about test findings must have something more than personal view and clinical judgment as proof of support for an opinion. Courtrooms like simple "Yes" and "No" answers and initially the valid/invalid distinction that SVT/PVT measures seemed to provide appeared to be tailor-made for forensic evaluations, especially when the malingering term could be used. But the opposite line of questioning and logic is just as problematic. As discussed in this review, the potential for a true malingerer to "game" the testing (pass the SVT/PVT measures but perform poorly on other tests feigning cognitive impairment) is likely occurring (Horwitz and McCaffrey, 2006). If an attorney

wanted believability as the goal for SVT/PVT findings, all that is needed is a little different line of questioning:

ATTORNEY: Dr. XX –"On the _ _ _ performance ***Validity*** test you administered, the examinee ***PASSED*** that measure, correct?

Psychologist: Yes

ATTORNEY: Valid performance on a SVT/PVT test is used to indicate believability of symptoms and test performance, is that not correct?

Psychologist: Yes

ATTORNEY: Consider this a hypothetical question and please answer accordingly. If someone is honestly telling the truth and generates valid test results, those test results should be believed, correct?

Psychologist: YES

Almost an identical line of questioning, but now the SVT/PVT finding is being used as proof of test result veracity with the implication that the individual is honestly performing. In these two scenarios, regardless of whether the test results are valid or not, the way in which the simplified pass versus fail distinction is made becomes problematic.

From an attorney's perspective, the veracity of test results and client statements become critical to defending, prosecuting, or representing a client. Furthermore, research to back one's position becomes critical in supporting or defending a position. These two factors and the demand placed on psychologists to utilize psychological and neuropsychological assessment data in the courtroom sparked a dramatic increase of published articles on the topic of 'forensic psychology'. As shown in Figure 5, derived from a research of the National Library of Medicine (pubmed.gov) using the terms "forensic" and "psychology" this line of research has steadily increased and more than doubled in the last 15 years. Note published research on this topic was relatively flat through the latter part of the twentieth century but increased exponentially during the twenty-first. This should also inform the reader as to the relatively beginning stage of the field.

FIGURE 5 From a National Library of Medicine search (pubmed.gov) that examined the frequency of publications with the key words of "forensic" and "psychology" (7,115 articles). Note the dramatic increase over the last 15 years. If the terms "forensic" "psychological" and "assessment" are entered in, there are only 551 articles. Further restricting the search to "forensic" "neuropsychological assessment" (355) or "forensic" "neuropsychology" (140 articles) even fewer articles are referenced. The figure graphically depicts the relative newness of the field in terms of published studies.

A research design problem with forensic-based SVT/PVT studies is the lack of forensic practitioners being independent in the classification and diagnosis of the examinee. In evidence-based medicine, to reach a Level I quality of evidence designation (see Neurology, 2011), to craft the highest level and most rigorous research design, investigators must be independent, blind to diagnosis and outcome, and some element of randomization must be present. As reviewed by Bigler (Bigler, 2006; Bigler et al., 2009), there are no forensic studies in the neuropsychological literature that could be considered Level I. As such, there are NO SVT/PVT studies that meet the highest levels of independence and research rigor. As already stated, there are no independent SVT/PVT studies that have been supported by external government agencies or private research foundations. Some forensic studies have pooled data and have impressive sample sizes, but large sample sizes do not overcome flaws in research design. As suggested by the examples of the attorney questions raised above and knowing that the first step prior to a psychologist examining a forensic case is the attorney or court referral, how has the forensic

literature on SVT/PVT measures handled this bias? This is essentially an unanswerable question at this stage as it has not been specifically addressed.

Some forensic SVT/PVT studies consider their design to be a “known groups design”. Having a truly known group is a bench-mark for evidence-based proof to measure diagnostic accuracy and classification. However, the limitation as already pointed out, is that the “known” group was never independently identified if the study is a retrospective, case review from a forensic practice and attorney or insurance referrals. Essentially all such SVT/PVT studies to date are quasi known group, crafted in a post-hoc fashion. As discussed in the next section, there is a lot written about SVT/PVT studies, but careful reading will show the lack of blindness and independence of investigators in such studies which, in turn, limit the generalizability of the findings. Individual forensic practitioners from the private sector doing research and publishing do not have any independent research oversight of their work product, do not have institutional review board issues that must be addressed, and, by the nature of the data collection, the research is funded by remuneration for being forensically involved in the case. Furthermore, careful reading of some of these investigations shows that by performing retrospective chart or case reviews, there will be no missing data. Large-scale prospective studies are very difficult to conduct even with substantial funding, where all data points are met.

The context of the forensic setting and nature of the referral may also represent another element unexplored in the forensic literature involving SVT/PVT findings (Murrie et al., 2013). In a personal injury forensic case there are typically two designated psychological/neuropsychological experts. For the plaintiff filing the case the plaintiff-retained psychologist/neuropsychologist may be perceived by the examinee as something more akin to a treating clinician and therefore more comfortable with the assessment. Oppositely, the psychologist/neuropsychologist retained by the defense may be perceived as an adversary and how this may engender more illness or sickness behavior, emotional or test/performance anxiety, has not been explored in any in-depth fashion. In personal litigation and most criminal cases, despite there being two retained experts, most of the SVT/PVT literature comes from data generated by only one of the evaluations. There are no large-scale studies that report how consistent or inconsistent test results are between the plaintiff-retained and defense-retained or prosecution-retained evaluations on SVT/PVT findings. This could be an important distinction because SVT/PVT data coming from the side that would examine the individual during the

second round of testing will know what tests were originally administered and likely select alternate SVT/PVT measures. How these factors might influence outcome is unknown. Referral and recruitment biases are known to influence neuropsychological outcome studies (Duquin et al., 2008; Paganini-Hill et al., 2013; Scotland et al., 2009). Referral and recruitment bias should be a focus of SVT/PVT studies as well.

Finally, because of the forensic implications and adversarial nature of this process, a "gotcha" approach occurs in some settings, placing some forensic psychologists on a mission to find the examinee with non-credible performance. Administering an excessive number of SVT/PVT measures in the hope of finding one below the cut-point to raise the suspicion of non-credibility is an example of this approach. The blurred roles that the forensic examiner and researcher have provide unique circumstances for unintended biases to cloud the clinical picture as to the meaning of SVT/PVT findings. Because so much of the SVT/PVT literature has come out of forensic settings, there are many potential challenges to the interpretation of SVT/PVT findings because of the above mentioned problems with research design.

SVT/PVT Science

What does the term "SVT/PVT science" mean? Within forensic SVT/PVT publications, the term "SVT/PVT science" sometimes will be mentioned in a statement like "…. SVT (or PVT) science has demonstrated …." Such a statement is followed by some presumed factual comment about the ability of SVT/PVT measures to identify valid and invalid test performance, implying scientific assurance in the accuracy of classification. Indisputably, scientific methods have been applied to SVT/PVT research. The dose-response effects of failing a SVT/PVT as described in the introduction have been independently replicated by numerous research studies using a variety of study designs. But why refer to this as SVT/PVT science? Is it not just SVT/PVT research at this point in time? The use of the term "science" implies some precision and finality to what may be concluded from the research. As reviewed herein, SVT/PVT research has a relatively short history, mostly spanning only the last three decades, with numerous limitations as already pointed out. Simply from that fact alone, there is much more research to be done, a seemingly basic requirement before the "science" label should be attached.

Some of the wording on how SVT/PVT measures are used has legal implications. There is an issue with admissibility of test findings in the court, often referred to as Daubert or Fry

hearings and avoidance of the junk science label (Samet and Burke, 2001). For admissibility there has to be an accepted scientific basis for a test finding. Some SVT/PVT measures have been legally challenged as to their admissibility in both civil and criminal courts and even excluded. Clearly there is good research and a liberal definition of science would apply to using the term. But to simply refer to this as "SVT/PVT science" implies that more is known about SVT/PVT effects than what may be the case.

Maybe the best argument why the "SVT/PVT science" term may not be the best descriptor comes from restricted conclusions made when the SVT/PVT score is below the cut-point, regardless of how far below. Collectively, this literature often makes the general conclusion to mark all test results invalid and to not interpret the findings when a SVT/PVT measure is failed. As already reviewed substantial numbers in clinical samples fail SVT/PVT measures; why? Merely attributing this to poor effort, lack of task engagement, malingering, and the like, is unsatisfactory and such explanations fall short of using a psychological or neuropsychological explanatory framework to more fully define the meaning of these SVT/PVT scores. If SVT/PVT researchers were seeking the broadest knowledge base to explain SVT/PVT findings, there needs to be a full explanation as to what SVT/PVT failure truly means, not just that such a measure was "failed". If the forensic end-point is merely to raise the specter of doubt (see Michaels, 2008), SVT/PVT findings can certainly be used to achieve that objective. However, as this review points out, high false-positive SVT/PVT rates for some neurological and neuropsychiatric disorders combined with not knowing precisely what SVT/PVT failure means or how to interpret such findings limits the current utility of SVT/PVT measures.

As explained in the Bigler (2012, 2014) reviews, there is a cognitive neuroscience to "effort" that includes both conscious and non-conscious perceptions and the valence of what may be perceived as a reward, both intrinsic and extrinsic, for completing a task (Bijleveld et al., 2014; Pessoa and Engelmann, 2010). Most SVT/PVT studies have not involved any aspect of neuroimaging or cognitive neuroscience to concomitantly explore what may be occurring at a neural level in response to a SVT/PVT task.

Finally, the absence of Level I evidence should be sufficient to question the "SVT/PVT science" designation. In studies that meet Level I design, there is also transparency and independence in funding. In SVT/PVT studies, how is funding for the study designated? Funding as a by-product of a clinical practice is not necessarily disclosed. Recently, journals have become

more stringent in their conflict of interest and disclosure statements, but journal articles from just a few years ago may have not have imposed any constraints on disclosure, including on those who profit from commercial sales of SVT/PVT measures (Bigler, 2006). There is nothing wrong with forensic practitioners publishing their findings, but full disclosure should be a requirement.

Can SVT/PVT Measures Diagnosis Malingering?

Bass and Halligan (2014) provide an excellent and detailed review on identifying factitious disorders and malingering. They make two important statements: "(1) A key challenge in any discussion of abnormal health-care-seeking behavior is the extent to which a person's reported symptoms are considered to be a product of choice, or psychopathology beyond volitional control, or perhaps both, and (2) Clinical skills alone are not typically sufficient for diagnosis or to detect malingering" (p. 1422). They also emphasize that in addition to medical diagnostic decision making that there needs to be "an increased appreciation of the contribution of non-medical factors and a greater awareness of the conceptual and clinical findings from social neuroscience, occupational health, and clinical psychology" (p. 1422).

Bass and Halligan (2014) put forth a model for differentiating factitious disorders and malingering that has direct relevance to SVT/PVT issues (see Figure 6). In this model they discuss "compensation neurosis," a term that had its origins from the twentieth century neurosis-driven psychodynamic nomenclature. There is relevance to this issue because in our contemporary litigious society certain injuries may be compensable, setting the stage for secondary gain. Secondary gain, just like placebo effects, can be powerful incentives and sufficient to alter findings in psychological and neuropsychological testing. However, note that in this model the dimensions are showed to be continuously transitioning, for example from non-intentional to exaggeration to malingering, with no clearly defined boundaries. This has implications for why SVT/PVT measures use a rigid cut-point applied to all cases in all situations for the simple valid/invalid dichotomy.

As stated in the introduction, below-chance performance on a SVT/PVT appears to meet the intentionality criteria needed for the malingering classification and deception in the Bass and Halligan model. Note in this model that definite malingering is at the far upper right corner within the box labelled deception. As an example of how a SVT/PVT finding could be used in this context, an individual who reportedly sustained a mild concussion wherein there is no

documented loss of consciousness and no retro- or anterograde amnesia but who exhibits profound memory deficits on neuropsychological testing and fails a 50-item forced choice SVT/PVT memory measure with a score of 7/50 is most likely malingering.

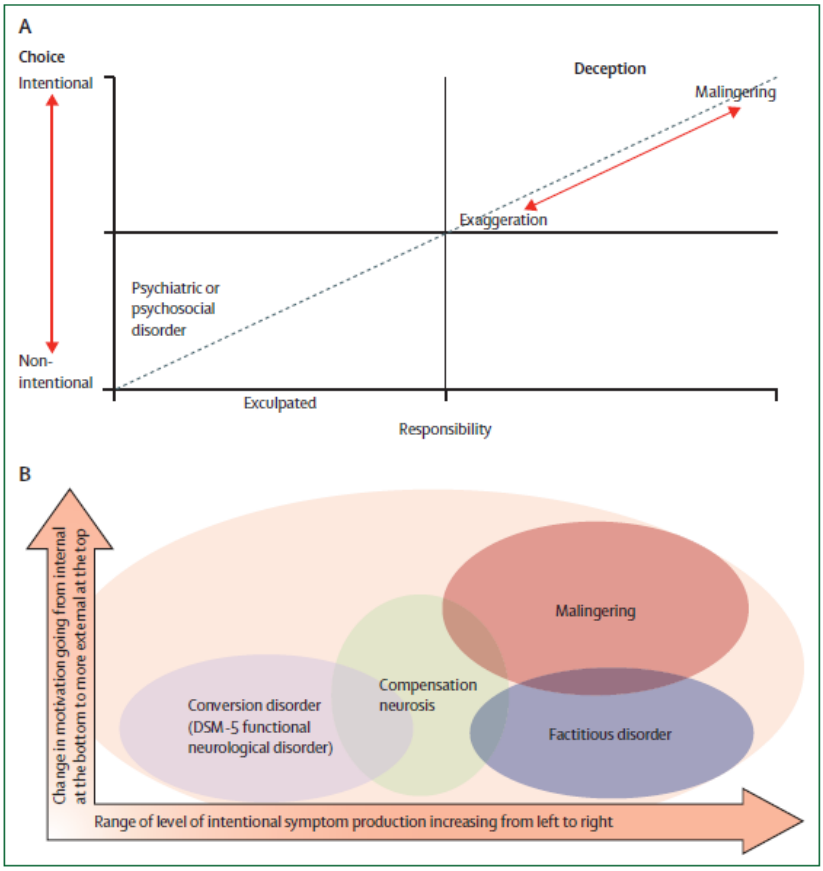

***Figure*: Two models of illness deception (A)[8] and compensation neurosis (B)[83]**
Reproduced by permission of Sage Publications (A) and American Psychiatric Press (B). Diagrams show the potential roles of patient choice, intentions, and motivation in symptom production and, ultimately, diagnosis. DSM-5–Diagnostic and Statistical Manual of Mental Disorders, fifth edition.

Figure 6 Two models of illness deception (A) and compensation (B)
Reproduced by permission of Sage Publications (A) and American Psychiatric Press (B). Diagrams show the potential roles of patient choice, intentions, and motivation in symptom production and, ultimately, diagnosis. DSM-5—Diagnostic and Statistical Manual of Mental Disorders, fifth edition.
SOURCE: Bass and Halligan (2014)

Below cut-score but above chance SVT/PVT performance does not provide unequivocal support for intentionality or a feigned response. Returning to the diagram in Figure 6, as one moves toward the bona fide psychiatric spectrum, conscious intentionality is no longer inferred,

but non-conscious intentionality seems a rather specious concept to objectively identify. Psychology and neuropsychology, although recognizing the importance of detecting and reporting malingering when thought to be present, have no uniform standards for diagnostic assurance that malingering is the explanation (Slick et al., 2004).

Therefore, Slick and colleagues in the field of neuropsychology (see Slick et al., 1999), p. 551) specified various conditions that should be met before the malingering term is used:

1. Meeting an operational definition of malingering
2. specific, unambiguous, and reliable criteria that cover all possible sources of evidence (i.e., test-performance, observations, and collateral data)
3. specification of the relative importance of diagnostic criteria
4. specification of the nature and role of clinical judgment
5. specification of differential diagnoses and exclusionary criteria
6. specification of levels of diagnostic certainty

These criteria, often referred to as the Slick et al. criteria, were proposed because "… the process of diagnosing malingering remains difficult and largely idiosyncratic" (p. 545), a conclusion mirrored by Bass and Halligan (2014). The Slick et al. criteria provide reasonable and generally accepted data points that span psychometric, behavioral, and collateral information. The criteria do permit the utilization SVT/PVT findings and provide a rationale for diagnosing malingering in some individuals. However, in reviewing the criteria, SVT/PVT findings are relevant only to one of the six criteria. Accordingly, a SVT/PVT finding by itself, even a below chance finding, does not permit the automatic conclusion that the patient is malingering. Furthermore, since SVT/PVT performance can be passed by a malingerer, the test alone is not fool-proof (DenBoer and Hall, 2007).

What has been overlooked in much of the SVT/PVT literature is that the Slick et al. criteria establish three classifications: "possible, probable, or definite malingering" (p. 551), yet the most frequent classification used in the SVT/PVT literature is "malingering," without any qualification, as if it were definite. Furthermore, much of the literature relies on the SVT/PVT measure(s) and simply below cut-score performance as the criterion for "non-credible" test performance, resulting in unavoidable circularity as to how malingered test performance is

defined. Malingering and non-credible get equated, but there may be many factors that contribute to non-credible test performance other than malingering. In some of these studies the assessment of malingering is inexorably connected with a diagnosis of malingering despite the absence of any external validation other than the SVT/PVT measures to define malingering. A major error of circularity in logic occurs in these types of studies. The reader has only to carefully examine what independent criteria were necessary for making an accurate and independent diagnosis of malingering in these studies, and, for some, it will quickly become apparent that inappropriate or incomplete criteria were used for the diagnosis of malingering.

Ambiguity in malingering classification when SVT/PVT measures are used occurs with certain groups who may have low educational attainment or English as a second language. All of the SVT/PVT measures currently in use were initially developed in English. Some have now been standardized in other languages, but this is not universal across SVT/PVT measures. Lower levels of education and testing in English when English is the examinee's second language have been shown to be associated with higher rates of below cut-score performance in as many as 20 percent of the examinees (Strutt et al., 2012).

Even if there is SVT/PVT failure suggestive of malingering, what is the clinician to decide in the case of a neurological patient with an obvious deficit who appears to be intentionally falsifying SVT/PVT performance? Bianchini and colleagues (2003) report on three cases in the context of litigation in which all of the litigants met criteria for having sustained a moderate-to-severe TBI and all met criteria for malingered test performance. The obvious problem is that an examinee may have a legitimate and compensable injury or disorder regardless of whether or not they are malingering. Some may exaggerate their symptoms as a "cry for help" (Berry et al., 1996; Blank et al., 2014).

As will be discussed more fully later in the review, there is no set number of SVT/PVT measures that should be administered. Some advocate the administration of multiple SVT/PVT measures (Larrabee, 2014a); however, multiple SVT/PVT administrations during an assessment also changes the likelihood for a failed SVT/PVT finding and misclassification (Berthelson et al., 2013). Nonetheless, two or more below chance SVT/PVT measures would certainly support presence of malingering. As Bass and Halligan (2014) point out there are multiple factors that must be taken into consideration when making the diagnosis of malingering, and the issue of

how multiple SVT/PVT administrations to the same examinee relates to classification accuracy is unknown at this time.

One final point on SVT/PVT research and malingering is that many of the studies are simulation studies. The typical design is one in which healthy controls characteristically are assigned to different groups in an attempt to mimic some type of impairment (as an example, see Barhon et al., 2014). Although these studies provide the precision of laboratory control, it is difficult to generalize such findings to real-world cases.

What is Effort to Perform a SVT/PVT Task? A conspicuous limitation of any technique that assesses an individual's symptoms or abilities is that the examinee's responses are under his/her control. The assumption that the demand characteristics and/or that "eyes-on" examiner identification detects invalid test performance through observation alone does not have empirical support. The issues of clinical training and clinical judgment are complex and critically important in making diagnostic conclusions and recommendations about whether an impairment or disability is present, and this includes interpreting SVT/PVT findings (Lezak et al., 2012). As Millis (2009) states:

> All cognitive tests require that patients give their best effort 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 et al., 1999).

As already discussed, schizophrenia is a disorder that affects task-engagement and motivation (Avery et al., 2009; Fervaha et al., 2014; Lafargue and Franck, 2009). On

neuropsychological testing, individuals with schizophrenia typically perform poorly on measures of executive functioning like the Wisconsin Card-Sorting Test (WCST; Van der Does and Van den Bosch, 1992). Thus under standard conditions of test administration, individuals diagnosed with schizophrenia tend to perform significantly worse than age and otherwise demographically matched controls because of poor effort and task engagement. The study by Stevens and colleagues (2014) demonstrates the point that even individuals with schizophrenia who pass SVT/PVT measures display impairments on complex cognitive processing tasks (see Figure 2). However, can this presumed core deficit of mental effort in schizophrenia be modified? Monetary incentives improve performance, as do other training methods, when applied to performance on the WCST by patients with schizophrenia (Bellack et al., 1990; Green et al., 1990; Nisbet et al., 1996). What does this mean for the administration of the WCST to assess executive ability when performed under standard conditions (i.e., "do your best") in someone with schizophrenia, where the group typical response is to underperform? Should the assessment only be done under conditions of incentive so that an optimal level of performance can be obtained? In standard administration, the examinees only would have been instructed to "do their best" and put forth their "best effort" in doing the task. In fact, that was the instruction used the study by Stevens and colleagues, yet, as a group, patients with schizophrenia performed poorly on the WCST regardless of whether the SVT/PVT was passed or failed. Those who failed the SVT/PVT measure performed the worst.

No standardized psychological or neuropsychological testing is done with incentive. No SVT/PVT measure is administered with an incentive other than an exhortation to "do your best." So whether someone is performing optimally or not is never really tested because incentive versus nonincentive conditions are never examined to determine if someone is performing optimally.

What would a "failed" SVT/PVT finding mean in a SSA disability examination of an individual with a diagnosis of schizophrenia? If schizophrenia is associated with expectedly high SVT/PVT failure rates (Gorissen et al., 2005), especially when low IQ is present (Sullivan et al., 2014), why would not a failed SVT/PVT measure in someone with the diagnosis of schizophrenia be expected and acceptable? If the SVT/PVT measure is tapping an element of poor effort inherent to the schizophrenic disorder then SVT/PVT failure may actually be diagnostic. In this sense, the SVT/PVT is not acting as a pure validity measure but as a maker of

the disorder. Recall that in the Stevens investigation, an almost identical pattern of impairment is seen in the schizophrenia group that "passed" as in the "fail" group on the neuropsychological battery of tests administered, the difference is just in magnitude (see Figure 2). The schizophrenia "fail" group generally ranged from 0.5 to 1.0 S.D. difference below the "pass" group, but the overall pattern is the same. On some measures like the Vocabulary Test and reaction time tasks, there were minimal or no differences, so the pass/fail SVT/PVT distinction was irrelevant for those measures. Why not interpret this pattern of deficits in patients with schizophrenia by acknowledging that underperformance might have been present as indicated by below cut-score results on the SVT/PVT, but noting that this is consistent with what would be expected in someone with schizophrenia? Note that if the psychologist merely interpreted the WCST as being "impaired" in the cases reported by Stevens and colleagues (2014) study, the clinician would not be making an interpretive error because the pass or fail SVT/PVT finding is irrelevant for that distinction – both groups performed below acceptable normative standards. The only difference is magnitude.

Similarly, in the study by Suchy and colleagues, the patients with MS all performed below control normative standards, irrespective of SVT/PVT findings (see Figure 1). Individuals with chronic MS may experience a wide array of psychosocial, medical, and motivational challenges (Simmons, 2010). For those in the Suchy study who "failed" the SVT/PVT, what evidence is there that the memory assessment findings that placed them as a group about 1.5 standard deviations below the mean of the normative sample as shown in Figure 1 was not accurately reflecting their perceived functional limitations in memory from their chronic MS?

Instead of using a SVT/PVT, Fervaha and colleagues (2014) took a different approach with independently diagnosed patients with chronic schizophrenia, all of whom were on stable medication regimens and were assessed using a quality-of-life scale that addressed intrinsic motivation. The sample was large (N = 431), drawn from multisite academic and community medical centers that involved participants from a number of different states. Motivation and cognitive performance were "robustly" related, suggesting that test findings from cognitive assessment are "… not purely a measure of ability (e-pub)." This further substantiates that the high SVT/PVT failure rates found in examinees with schizophrenia and other neurological and neuropsychiatric disorders with similar underlying neuropathology likely occur because of core

motivational factors. As already stated, SVT/PVT measures may be directly assessing core features of schizophrenia.

In summary, given the complex relations between motivation and cognition and the generally high failure rates on SVT/PVT measures (Hunt et al., 2014; Morra et al., 2014), although not all patients with schizophrenia show this tendency (see Strauss et al., 2014), a SSA examiner will be confronted with an impossible task to ferret-out the true meaning of an above chance but below cut-score result on a SVT/PVT measure in someone with a diagnosis of schizophrenia.

As outlined in the Bigler (2014) review, there is a neurobiology of effort that has basically been ignored by SVT/PVT researchers. In some of the early SVT/PVT publications, there were statements inferring that the simplicity of the SVT/PVT task required "no effort" to perform. This is, of course, incorrect as shown by functional neuroimaging studies using SVT/PVT measures, which indicate characteristic activation of language, memory, and executive functioning brain regions to perform the task (Allen et al., 2011; Wu et al., 2010). The motivational systems of the brain, as well as memory and executive functioning networks, involve frontotemporolimbic and default mode networks of the brain. These are the "mental effort" systems of the brain, yet there is a paucity of neuroimaging or biomarker studies examining the relations of these brain regions and SVT/PVT findings. The importance of this becomes obvious when it is understood that these brain regions are the ones most consistently identified as abnormal in neuropsychiatric disorders, including schizophrenia and other neurological disorders where there is SVT/PVT failure.

Loring et al. (2011) directly explored the issue of mental effort by examining SVT/PVT performance during exposure to a benzodiazepine (lorazepam) in a randomized, double-blind, placebo controlled, crossover trial. As would be expected, since benzodiazepines affect error monitoring (Riba et al., 2005), SVT/PVT failures would occur during lorazepam administration. This is one of the few SVT/PVT studies to rigorously explore what these investigators label as "… potential latent variables" (p. 799) and their influence on SVT/PVT performance. If a disorder or disease disrupts the error monitoring or attentional networks as lorazepam did in the Loring study, then SVT/PVT failure may occur as a consequence of an underlying deficit.

SVT/PVT Terminology–How Should SVT Results Be Reported and Stated?

Pass/Fail and Valid/Invalid have been the most common terms used to describe SVT/PVT scores. However, as already argued, should a score just below the cut-point really be called a failure or invalidity? How should these findings be reported without pejorative connotation? There is a long list of terms used to describe below cut-point SVT/PVT scores including: non-credible, disingenuous, non-believable, feigned, faked, falsified, fabricated, exaggerated, embellished, manufactured, and malingered. When and under what circumstances should any of these terms be in a report?

Bigler (2012) presented two imaging cases with clear-cut pathology from TBI that damaged attentional and default mode networks. The individuals in both cases missed the cut-points on SVT/PVT measures by just a few points, thus, far above chance. The definitive pathology disrupting attentional performance is likely a factor in the "failed" SVT/PVT performances, but should the results be reported as "failures"? How should the SVT/PVT performances be explained to the patients, the referring neurologists, and potentially a SSA examiner? Psychological and neuropsychological evaluations become memorialized in the medical and disability records. The conundrum in reporting a "failed" SVT/PVT in someone with a legitimate disorder is characterized in the following hypothetical case report:

> Patient XXX sustained a severe TBI with an initial Glasgow Coma Scale score of 3 and day-of-injury CT imaging showing extensive frontal and parietal contusions. Follow-up imaging demonstrated extensive frontal pathology, and behaviorally on the inpatient rehabilitation unit he displayed poor motivation, emotional dysregulation and poor task engagement consistent with frontal lobe damage. Unfortunately, he "FAILED" his SVT/PVT measures, and therefore I can't interpret any of his neuropsychological test findings.

The pejorative quality to "failure" or "failed" (or any derivative thereof or associated term) in cases like this is obvious. Since the SVT/PVT results in such cases are most likely not failures, but explained by the brain pathology, how should they be described? SVT/PVT proponents have skirted this important issue. Walter and colleagues (2014) conducted a study

that examined severity of dementia and performance on a SVT/PVT measure and found that 20 percent of the participants in the group with severe dementia "failed" the SVT/PVT measure. If one-fifth of those with a legitimate dementia "fail" the SVT/PVT because of their disease, what justification does a psychologist have to call something a "failure" when it is part of an expected pattern of results that may occur as a result of the disorder? Use of the current SVT/PVT terminology of "failed" or "invalid" runs the risk of legitimate cases being discounted or misconstrued or their significance being diminished in some way. If a rigid, simple binary fail/pass or valid/invalid distinction is adhered to, then all legitimate cases who perform below the cut-point will be labeled with "failed" or "invalid" performance. Some of the problem with classification could be addressed if more flexible algorithms were established with variable cut-points based on the clinical history, medical information, and, potentially, neuroimaging findings.

Is Deception Part of the Effect for Performance Validity Testing of Cognition

Should examinees be informed about the nature of the tests they are taking? Should they be informed if they perform below a cut score and given an opportunity to improve? In other words, when it comes time to administer a SVT/PVT, should the examinee be told the true nature of the test? This is not a new debate for psychology. Any test environment is a study in simulation of real-world settings. Data from an intelligence test is but a simulation of real-world questions that inferentially "measure" the construct defined as an intellectual ability. The better the simulation of real-world circumstances, the more reliable is the inference. All that is typically done in the contemporary psychological or neuropsychological assessment is to inform the examinee that various cognitive abilities will be tested and that they should try their "best" at all times during the assessment, but no specifics are given about when and which SVT/PVT measure(s) is(are) administered. When I examine a patient, in addition to what was just stated, I tell them that during the assessment process tasks will be given that will provide information on "how hard they are trying." I tell them, "Always try your best but let me know if you feel fatigued or need a break. We need you to be trying as best you can throughout the testing that we do today."

Social psychology has hotly debated the importance of simulation designs and deception since the classic Milgram experiments on obedience in the 1960s (Milgram, 1963). The basic

simulation in SVT/PVT tasks is that examinees believe that they are indeed performing a task that is genuinely assessing a cognitive or emotional symptom, not just validity. The SVT/PVT task is administered among all of the other tests, never separately identified as a validity indicator. As previously outlined, the common theme for PVTs has been the ease of the task, wherein there is almost universal success of "passing" the PVT measure with few to no errors in healthy controls. This facet of PVT test development and standardization cannot be disputed. Because the task is so basic and easy the argument is made that the examinee does not need to be informed. But does this practice meet informed consent requirements?

As discussed, certain psychiatric and neurological conditions directly influence test engagement, but if the patient is not informed that their ability to perform is being questioned, they may not exhibit motivation to perform at optimal levels. Does this matter? The study by Suchy and colleagues (2012) in patients with multiple sclerosis indicated that for some who performed below SVT/PVT cut scores that confrontation with their PVT/SVT failure allowed them to pass a re-administration of the SVT/PVT. This in turn yielded what were deemed to be valid test results. In other words, from a technical SVT/PVT binary classification they went from invalid to valid performance, when informed about the nature of the SVT/PVT task and asked to perform the task again. This method may indeed be an appropriate use of an external SVT/PVT. If such a measure is given and failed, improved performance on a second try following confrontation is proof that suboptimal performance occurred on the first trial.

Unfortunately, there is very little research on this topic. Johnson and colleagues (1998) and Johnson and Lesniak-Karpiak (1997) used undergraduate students in a simulation study who were instructed to feign cognitive performance on standardized testing. Warning that that feigning could be detected resulted in test performance similar to controls compared to those who received no warning. Contrived studies that use undergraduate participants are fraught with design and generalizability problems, so little can be concluded from these kinds of studies. Conversely, the findings of Youngjohn, Lees-Haley and Binder (1999) are summarized by the title of their study, "Warning malingerers produces more sophisticated malingering." Nonetheless, Schenk and Sullivan (2010)(p. 752) conclude that "that using a carefully designed warning may be useful for reducing the rate of malingering," and if the intent of psychological/neuropsychological assessment is to obtain a valid assessment, then why not inform? That information about the nature of SVT/PVT measures may merely lead to more

sophisticated malingering is certainly a genuine concern (Franzen and Martin, 1996), but informing the examinee about what the procedures are that they are receiving as part of an evaluation also is a feature of informed consent.

This would be less of an issue if each test had embedded within it a specific measure of validity. As currently used, external SVT/PVT measures pose another validity challenge; such measure are subject to "coaching" if they are identified by the examinee (Rose et al., 1998; Suhr, 2002; Suhr and Gunstad, 2000). If warned, one could be "coached" to perform well on the tasks that appear "easy" and not so well on other test items that appear to be a more genuine measure of cognition. DenBoer and Hall (2007) demonstrated that "brain injury simulators" can be coached to pass SVT/PVT measures but perform poorly on other neuropsychological measures. Obviously, if someone had malicious intent they could be coached to malinger (Tan et al., 2002; Trueblood and Schmidt, 1993), defeating the purpose of having an external measure. Indeed there are webpages that examinees may query prior to evaluation (Bauer and McCaffrey, 2006)

Returning to the issue of deception in SVT/PVT use, some of the deception effect of the SVT/PVT may be in the manner in which it is presented to the examinee. For example, Morel and Marshman (2008) discuss the development of the Morel Emotional Numbing Test (MENT), designed as a forced choice SVT to detect genuineness of symptoms associated with post-traumatic stress disorder (PTSD). The MENT employs a forced choice paradigm for identifying facial expressions. The assumption about the MENT is that common facial expression is universally recognized, typically with minimal effort. However, in the instructions for administering the MENT a potential cognitive set is subtly reinforced by indicating to the examinee that some people with PTSD ‘‘may’’ have difficulty recognizing facial expressions. As already stated and shown by Darwin (1872) almost 150 years ago, basic facial emotions and expression are commonly recognizable across all peoples and cultures, so the MENT is easily passed, even by those with PTSD. This minimal deception however, is enough to result in individuals who may be exaggerating or feigning PTSD symptoms to fail to identify common facial expressions (Morel and Shepherd, 2008). However, when Pedersen and colleagues (2014) removed reference to PTSD in the instructions to individuals seen in a VA Center with either a diagnosis of PTSD or self-reported PTSD, error scores and failure rates significantly decreased leading them to conclude that "… deceptive language is seemingly necessary for eliciting PTSD symptom exaggeration with the MENT" (p. 603).

The role that deception may play in SVT/PVT administration and performance has not been extensively investigated.

Legitimate Impairments and SVT/PVT Failure

In addition to discussing the MENT, Morel and Marshman (2008) provide an excellent review outlining the requirements of what should be critical elements of SVT/PVT measures which include the following: "(1) sensitivity and specificity, (2) face validity, (3) measurement of symptoms relevant to the claimed dysfunction, (4) normative data sufficient to meet scientific and legal standards, and validation with known groups (i.e., real malingerers), (5) resistance to faking or coaching, (6) ease of administration and interpretation, and (7) ongoing research." As recognized by Morel in the development of the MENT, issues of sensitivity and specificity are influenced by presence of disorders where legitimate impairment may be tapped by certain conditions. Since the MENT is based on a forced choice facial recognition paradigm, in the norming of the test it was important to include subgroups of patients who "… have genuine impairment in identifying facial expressions" (p. 1545), like patients with schizophrenia or autism. "For this reason, institutionalized patients with schizophrenia were included in the normative sample of the MENT to provide a baseline for legitimately poor performance" (p.1544).

Essentially all commercially available SVT/PVT measures have incorporated into their normative sample neuropsychiatric and neurological participants to address the issues labeled by Morel and Marshman as "genuine impairment" and "legitimately poor performance" that could result in SVT/PVT errors.

Unfortunately, not much can be said in this section. There are no large scale studies that address SVT/PVT failure in individuals with legitimate impairments because none of the SVT/PVT measures stratify groups with potential "genuine impairment" in any systematic or comprehensive fashion. As already reviewed, a high percentage of patients with schizophrenia fail SVT/PVT measures. Patients with various types of dementia fail, leading Bortnik, Horner and Bachman (2013) to conclude "the need for caution in interpreting effort measure performance in dementia samples due to the fact that despite their best effort, many patients with dementia fail effort measures and are at risk for being misclassified" (p. E-pub). Some SVT/PVT test publishers have now come out with a "profile" analysis for SVT/PVT test findings to reduce

false positive error rates, but the implication of this is obvious. The purported SVT/PVT measure that supposedly requires no mental effort and is acting solely as a validity indicator, is indeed tapping cognitive functions that are an aspect of the neurological or neuropsychiatric disorder. So why call it a SVT/PVT in the first place unless better classification accuracy can be obtained?

What patient groups with "legitimate reasons" perform poorly on SVT/PVTs and can they be differentiated? This is mostly an unanswered question because of the absence of any large scale, integrated, and independent investigation seeking an answer. Such a study would need to compare multiple SVT/PVT measures across all of the major neuropsychiatric and neurological disorders for which disability determination may be a reason for consultation and assessment. As already shown in this review, when assessing various groups with legitimate disorders, false-positive SVT/PVT identification rates occur with a range from 5 to 30 percent. These are simply unacceptable rates for SSA to utilize a specific SVT/PVT rule in disability determination.

In an examination of PNES patients who failed SVT/PVT measures (see Williamson et al., 2012), these investigators found that presence of SVT/PVT failure "… in 91 participants with PNES was strongly associated with reported abuse but, contrary to expectations, was not associated with the presence of financial incentives or severity of reported psychopathology" (p. 588). Similarly, within a military setting, Clark and colleagues (2014) found "…a greater prevalence of traumatic brain injury (TBI), Post-Traumatic Stress Disorder (PTSD), co-morbid TBI/PTSD, and other Axis I diagnoses, was observed among participants with poor effort" (p.802). The implication may be that the findings are not valid, but it may also be that these disorders may also be uniquely affecting test engagement, motivation, and attention to perform a SVT/PVT task.

In reading the forensic SVT/PVT literature, much of the conclusions focus solely on the secondary gain issue, although it is rarely scrutinized as a specific variable, just as some general condition (e.g., the examinee was in litigation). Another military study by Nave and colleagues (2014) tested this hypothesis by assuming that military personnel coming before a medical evaluation board (MEB) would be associated with a higher level of dissimulation because of secondary gain. They found that MEB review was not associated with any higher rate of SVT/PVT failure. What they did observe was that on a self-report personality inventory there was a tendency for over focus on negative psychiatric symptoms in those coming to MEB

review. Potentially, this could be a feature of the "cry-for-help" response, as previously mentioned. Poor performance on SVT/PVT measures may also be a part of the spectrum of "catastrophizing" after what may appear to be a minor injury (Miro et al., 2008). In this sense, SVT/PVT failure may be tapping premorbid neuropsychiatric risk factors (Lange et al., 2013; Larson et al., 2013). In these scenarios of SVT/PVT failure, it may be that the neurocognitive test results are less than optimal and potentially non-credible, but the reason may be that the patient is attempting to emphasize their impairments (exaggerate?), which still may be present, although masked by their embellished performance.

Mooney et al. (2005) in a university-based rehabilitation setting found that a group of 67 adults who had documented mild TBI but poor recovery was characterized by "… depression, pain and symptom invalidity" (p. 975). Some researchers have attempted to address the issues of pain, pain and depression and just depression on SVT/PVT performance (Alfano, 2006; Etherton et al., 2005), but the issues are complex. The border zones, as shown in Figure 6, become very blurred if one tries to use only a SVT/PVT to distinguish between bona fide illness and somatization when dealing with neuropsychiatric disorders and even malingering.

How Many SVT/PVT Measures Should Be Given?

The irony of the question of how many SVT/PVT measures should be given is missed by proponents of external SVT/PVT methods. The real question for psychology and neuropsychology is why are psychological/neuropsychological tests that require an external measure of validity (i.e., that do not have their own internal validity metric) still being administered? Do not studies that show high SVT/PVT failure rates merely point to the potential inadequacies of contemporary symptom and cognitive assessment metrics? Studies can be found that support the administration of multiple SVT/PVT measures to increase the detection of non-credible test performance and likewise argue that such multiple SVT/PVT measures are neither redundant nor interdependent (Larrabee, 2014a). However, as demonstrated by Bilder, Sugar, and Hellemann (2015) the independence of these measures cannot be supported (see also Berthelson et al., 2013).

Young (2014) discusses the majority of SVT/PVT tests in current use and the ethical implementation of their use and interpretation. Although a "top 10" list of most commonly used SVT/PVT measures could be established based on reported usage(see Fox, 2011; Williams,

2011), of the more than 40 measures that could potentially be used, as listed by Young, there has never been a comprehensive comparison across SVT/PVT measures . Although there are studies that compare some SVT/PVT measures, they typically examine head-to-head just three to eight such measures or procedures and only within a restricted sample associated with a particular disorder such as TBI (e.g. see Bashem et al., 2014). How SVT/PVT measures compare to one another and within and across different neurological and neuropsychiatric disorders is unknown.

If multiple SVT/PVT measures are administered, in what order should they be administered? The answer to this question also is unknown. As Bilder and colleagues (2015) have shown, the assumption that test performance on different SVT/PVT measures is not independent implies that the order of SVT/PVT administration will influence SVT/PVT performance. Years ago, Guilmette and colleagues (1996) demonstrated that there was an order effect of when a forced choice PVT measure was administered, where differences occurred depending on whether it was administered first or last. For example, a comprehensive neuropsychological assessment may take a day, sometimes split over two days. Do SVT/PVT measures tap fatigue at the end of a testing situation? If so, the number and order of SVT/PVT measures and when they are administered will be important. Decades of cognitive psychology research supports the influence of order effects with differing sensitivities depending on order of test administration. How testing fatigue influences SVT/PVT scores is not known. However, Mosti and colleagues (2014) examined sleep deficits and SVT/PVT scores in mild TBI patients, finding worse sleep quality associated with lower SVT/PVT performance on some measures. The implication is that vigilance and fatigue may be important for SVT/PVT results. Only a handful of SVT/PVT studies have empirically examined such things as distractibility and cognitive load and their potential influence on SVT/PVT measures (Barhon et al., 2014; Batt et al., 2008; Bowden et al., 2006). Additionally the order effects are not just the order of the SVT/PVT measures but the order of the main testing protocol and where and when the SVT/PVT measure is inserted. For example, a typical full administration of an intelligence test may take more than an hour of concentration and task involvement on the part of the examinee. Should the SVT/PVT be given before, in the middle, or at the end of intelligence testing? Simple motor and sensory-perceptual testing is less cognitively demanding. Is there a difference if the SVT/PVT is administered just before or just after a demanding task as compared to a less demanding task? These are all unknown clinical circumstances.

As mentioned in the introduction, computerized assessment methods that evaluate cognitive functioning are beginning to explore and use computer technology so that a PVT measure is actually integrated within the testing, but this approach to embedded SVT/PVT measures is just beginning (Brooks, Sherman, & Iverson, 2014; Roebuck-Spencer, Vincent, Gilliland, Johnson, & Cooper, 2013).

Which SVT/PVT Measure Should Be Administered? When During an Assessment Should They Be Administered?

As previously stated, there is no professional agreement on which measure is best all around, but even if there were a "best" omnibus SVT/PVT measure, it is very likely that certain diseases or disorders may disproportionately influence SVT/PVT performance. Again without a "gold standard" SVT/PVT measure, how would SSA make a decision on what may be considered the most reliable method or procedure to use?

Some of the work that attempts to show which measure may be most sensitive in detecting genuine invalid performance is highly problematic for making any general clinical recommendation. Research that uses analogue simulation-based methods can mostly be discounted and will not be discussed because these investigations often use student volunteers to simulate a disorder or even malingering. Such studies do not use clinical populations and real-world scenarios, although they may have a reference group with a particular disorder. As shown by Swihart and colleagues (2008) when tested in real-world mixed neurological and neuropsychiatric cases, the level and accuracy of detection is far from what is reported in simulation studies. Some of the work can also be discounted because it comes from authors of SVT/PVT measures that have a vested interest in outcome of the study. For the few comparative studies that would meet appropriate standards, often one SVT/PVT measure yields higher rates of failure. Some would claim that the measure with the highest failure rate is the better measure because it appears better at detecting invalid performance. But it may also be that the PVT measures with the highest failure rates are the ones tapping the cognitive domain where legitimate impairment is present. There are no studies that can be cited at this time that comprehensively address these issues.

The typical neuropsychological assessment, if it is to be comprehensive will assess sensory-perceptual, motor, spatial, and visuospatial ability, as well as memory and executive

functioning, intellectual and academic levels, and some aspect of personality/emotional functioning. However, as already stated, the preponderance of SVT/PVT measures are memory based. This raises all kinds of confounds when it comes to the type of disability that one has. For example, one may have a completely legitimate motor impairment and compensable disability. There are PVT measures for grip strength,(King, 1998; Simonsen, 1996; Smith et al., 1989) where expected force matches clinical presentation, that could be passed, but if the individual failed a cognitive PVT measure, what bearing would that have on the claim, if the claim were focused on motor impairment? It seems only reasonable that if a SVT/PVT measure is to be used it should be specific to a particular domain of cognitive, emotional, or behavioral functioning. Unfortunately, this is not the approach that developers of SVT/PVT methods have taken.

The order effects as described above are unresolved and without that resolution there can be no resolution on which SVT/PVT to administer, when in the assessment protocol it should be administered, and for what disorder.

SVT/PVT Fundamental Assumptions

There are fundamental requirements that should be met before a SVT or PVT task is administered. These fundamentals apply to any psychological, cognitive or behavioral assessment. Traditional validity indicators, whether SVT or PVT measures are administered or not, should always begin with an assumption that the examinee has the actual ability to perform the task. In other words the examinee has the mental capacity and possesses sufficient sensory processing and motor or verbal output to adequately respond. If hearing, reading, and/or visual processing are the focus of the disability in question, then a SVT/PVT given in the mode of disability may not be valid. On the other hand, there are SVT/PVT techniques that assist in defining whether a sensory deficit may be genuine or not (see previous discussion).

Another potential invalidating circumstance may occur when a SVT/PVT measure is given in a language other than what the SVT/PVT measure was standardized in. If sufficient mastery of the second language can be demonstrated, the SVT/PVT likely can be validly administered, although there are only a few studies that have examined this point (Burton et al., 2012; Vilar-Lopez et al., 2007). Education is a major factor as well, where individuals who are less well educated have higher failure rates (Loring et al., 2005; Webb et al., 2012). Violations of any of the above standards represent the basis for declaring invalid any SVT or PVT

administered, which would also call into question the validity of any standard psychological or neuropsychological measure.

Some symptoms, in part may be influenced by culture and language (Dere et al., 2013; Manly, 2008; Rivera Mindt et al., 2010). Indeed, both DSM-IV and DSM-5 contain glossaries of "Cultural Concepts of Distress." The potential cultural influences on SVT/PVT measures are unexplored.

RESPONSES TO SPECIFIC COMMITTEE QUESTIONS

In whom are PVTs and SVTs useful for informing disability determinations? In what way?

Administration of a SVT/PVT task or tasks is part of current practice standards endorsed by professional psychological and neuropsychological organizations. Fortunately, the majority of examinees assessed with psychological and neuropsychological tests pass SVT/PVT measures. A passed SVT/PVT is probably the best method currently to support the validity of the other test results. Unfortunately, there is no universal agreement on which SVT/PVT measure or measures should be administered and for what conditions.

The real issue with SVT/PVT use is what false positive rate is SSA willing to tolerate? Also, how should above chance and near miss, just below cut-score, SVT/PVT findings be interpreted and what lexicon should be used to describe such findings?

The study by Marcopulos and colleagues (2014) highlights these issues because the participants in that study would be typical of individuals potentially seeking SSA assistance at some point. Marcopulos and colleagues examined SVT/PVT findings in 436 consecutive referrals to a state psychiatric hospital examining both civilly committed and forensic cases. SVT/PVT failure was observed in 18.6 percent of the cases and approximately 8 percent had no indication for secondary gain. Individuals in this study with no apparent secondary gain tended to be older, female, Caucasian, and civilly committed patients in which bona fide disorders of cognitive impairment, behavioral issues, and inattention were considered the reason for PVT/SVT failure. Taking other considerations into account, a 10 to 15 percent SVT/PVT failure rate for reasons that could be considered part of their disability occurred in this institutionalized sample. These were only cases referred and not an assessment based on admission or diagnosis, which would have likely yielded higher results as evidenced by in the study by Stevens and

colleagues (2014) that reported a 26 percent SVT/PVT failure rate in a schizophrenic sample. Nonetheless, a potential false-positive rate of 10 to 30 percent in individuals with chronic mental health problems is an unacceptable rate to recommend any standard SVT/PVT procedure.

Turning to the private sector and a much larger scope of medical and neurological conditions for which psychological and neuropsychological assessments have been performed, the potential false positive rate is just as problematic. Frazier and colleagues (2007) found in a large hospital clinical setting, up to 30 percent of a mixed neuropsychiatric group referred for evaluation failed SVT/PVT tasks. Focusing on chronic epilepsy, a disorder with high rates of SSA disability allowance, Keary and colleagues examined 404 patients with intractable epilepsy and found 53 (13.1%) had questionable to invalid SVT/PVT scores. All had independently identified intractable epilepsy so their SVT/PVT "failure" would not be a determining factor related to the presence of disability. Further mitigating the clinical significance of the SVT/PVT failures, poor SVT/PVT performance in those with epilepsy was related to memory and intellectual abilities. The SVT/PVT performance was likely being influenced, at least in some, by the underlying neurological impairment and therefore SVT/PVT failure was false positive {see also \, 2005 #216}.

Chafetz, Prentkowski, and Rao (2011) found that approximately 40 percent of individuals in their sample of clinical referrals for Social Security disability determination failed SVT/PVT tasks. This study was not designed to further explore false positive rates, but given these findings and the failure rates just described for institutionalized psychiatric patients, general medical, and chronic epilepsy patients, a projection of 25 percent SVT/PVT failure rate or higher could easily be seen in those seeking SSA assistance.

If SSA utilized some criterion for SVT/PVT measures to be used in every evaluation, the SSA examiner reviewing a psychological or neuropsychological report with a passed SVT/PVT may not encounter any validity issues that need to be addressed, other than the test results may be accepted as valid. However, when a SVT/PVT failure has occurred, especially in patient groups where there are high rates of false positive SVT/PVT findings, how would a SSA examiner know that the individual has a legitimate disorder and the SVT/PVT finding was merely a false-positive? This would be entirely dependent on an elaborate discussion by the psychologist examiner to provide detailed commentary on the meaning of the SVT/PVT.

How/in what way do the results of PVTs or SVTs correlate with assessing functional limitations (such as limitations in a person's ability to do basic work activities, activities of daily living, social functioning, and concentration, persistence, or pace) due to an impairment?

In my opinion, this is completely an unanswerable question at this time. Almost all SVT/PVT research has examined symptom/problem complaints and cognitive functioning and has not examined activities of daily living (ADL), functional capacity, return-to-work, or issues of social functioning. In a National Library of Medicine search of all SVT/PVT/EF word combinations with the above 'functional limitation' term only one study reports on ADLs (see Cottingham et al., 2014), but that study used the discrepancy between level of reported cognitive impairment but intact ADLs as a marker of non-credible neuropsychological test performance.

Recalling that most of the SVT/PVT research has been generated by criminal and civil litigation cases and not SSA Disability [or if disability claimants were part of the study, most often they were not separately examined – exception to this is the work by Chafetz (Chafetz, 2011; Chafetz, 2012)], ADLs and related functional abilities have not been the focus of SVT/PVT investigations. Nonetheless, from the above discussion there will most likely be something in the range of a 10 to 40 percent SVT/PVT failure rate that the SSA would have to address. In an older study, Guilmette et al. (1994) examined 50 disability claimants referred by SSA for evaluation. One-fifth failed a forced choice PVT, with another 20 percent performing in an intermediate range. However, none of this work specifically addresses functional limitations. The inference from this would be that a claimant could indicate significant functional impairment and perform below a SVT/PVT cut-score, but do the SVT/PVT findings generalize to functional impairment and how would the legitimate case be differentiated from the non-credible?

What is the existing literature, if any, comparing the use of SVTs and PVTs in assessing the credibility of a claimant's statements about his or her symptoms (and their effects on his or her functioning) with a multi-domain assessment such as that currently used by SSA—i.e.,

- **The claimant's medical history, diagnosis, and prescribed treatment;**
- **The claimant's daily activities and efforts to work;**

- **Any other evidence showing how the claimant's impairment(s) and any related symptoms affect his or her ability to work (or, for a child, his or her ability to function compared to that of other children the same age who do not have impairments);**
- **Any observations about the claimant recorded by SSA claims representatives during interview (in person or by telephone).**

Current SVT/PVT literature does not specifically address these critical questions. There are no comparative studies that have examined the credibility of SSA claimants based on a variety of potential SVT/PVT measures that could be used. SVT/PVT studies have focused on the purported validity of the administered battery of psychological and neuropsychological tests and not on the "…credibility of a claimant's statements about his or her symptoms with a multi-domain assessment such as that currently used by SSA." A large scale, independent investigation prospectively examining SSA claimants with different diagnoses and levels of function would need to be undertaken using some of the most frequently administered SVT/PVT measures in use to answer this question.

Given the historical context in which SVTs were developed for forensic use in litigation settings, can they be adapted for use in disability determinations? Discuss the transferability of SVTs given the in evidence use/decision-making between fields (legal vs. mediated/negotiated).

As previously mentioned, most SVT/PVT measures have been examined in civil and criminal litigation cases and not in the SSA disability context. When SSA disability cases are identified, most often they are mixed in with civil and criminal cases and not separately examined. In a National Library of Medicine search using various key words for SSA and PVT/SVT measures there are less than 15 published studies addressing some of the above issues, and none use an independent sampling method, since all are from clinicians performing SSA evaluations and retrospectively examining their findings. If the question is "***Malingering***" then any external SVT/PVT that has a forced choice approach permits the application of the binomial statistic with below chance scores being the only direct method that assesses intentionality and choice to select the incorrect answer. The one caveat is that some patients with legitimate and compensable disorders, in a cry for help, may exaggerate or frankly malinger the level of their

cognitive deficit. Applying the Slick et al. criteria as outlined in the body of this review combined with SVT/PVT is the recommended approach currently used to diagnosis malingering.

How should one interpret scores/results in the "grey area" between clear failures (e.g., below chance scores) and clear passes on SVTs or PVTs? How many people fail completely vs. at the margins?

At this point, it becomes merely a clinical interpretation. As I have hopefully been able to elaborate in this review, the set cut-points for SVT/PVT "pass/fail" designations are problematic. Most research does not provide the kinds of details to calculate the numbers within the "grey area", but in the study by Locke and colleagues (2008) study, as previously discussed, provided sufficient information for this to be calculated. More than two-thirds of the approximately 20 percent of subjects who failed the SVT/PVT measure scored substantially above chance. As stated in this review, grey area failure is likely to be the most common occurrence with a high likelihood of not being a true failure. Such individuals would be misclassified if strict cut-scores rules were imposed.

When interpreting PVT or SVT failures, particularly in the "grey zone," are there factors aside from malingering or intentionally poor performance that may explain the results (e.g., stems from symptoms, fatigue, apathy)?

All of these factors may influence SVT/PVT performance. As shown in the review in TBI with typical frontal and temporal lobe damage being the norm, this could be a major factor. Chronic pain and how it affects testing, concentration, and task engagement has never systematically been explored. Greiffenstein and colleagues (2013) examined a group of individuals undergoing assessment for complex regional pain syndrome CRPS) and found that 75 percent "failed" at least one SVT/PVT. Johnson-Green and colleagues (Johnson-Greene et al., 2013) examined patients with fibromyalgia and found that 37 percent of the patients failed at least one SVT/PVT measure. Neither study was designed to ferret out the role of pain on SVT/PVT results, and since there is no independent biomarker to establish an objective measure of pain in a clinical or forensic setting, neither study could effectively differentiate between those with legitimate pain-related problems and those with embellishment or exaggeration of symptoms.

As reviewed above, apathy (lack of drive or motivation), which characterizes a core symptom in many patients with schizophrenia, probably relates to the high failure rates of SVT/PVT performance in that disorder. This may also be an explanation for high failure rates in samples of patients with epilepsy and numerous other neurological conditions. In both patients with schizophrenia and epilepsy, the number and type of medications may also influence SVT/PVT performance, but this has never been systematically examined. There are high rates of apathy in various degenerative disorders, chronic major depression, and neoplastic disease of the brain, especially when frontal or temporal lobe regions are involved, as well as in other bona fide neurological and medical conditions. There are no SVT/PVT studies where these factors have been systematically studied.

How does the current norming of SVTs and PVTs affect their usefulness in a variety of different populations (e.g., a diversity of race, ethnicity, culture, and educational or socioeconomic status)? Are there ways to resolve or mitigate the challenges posed by lack of norming for particular populations?

It should be a fairly straightforward conclusion from this review that none of these issues are adequately addressed.

Additional Questions

Does any credible literature exist that addresses variation in interpretation of SVT and PVT scores? In other words, a score is generated, but it appears there is room for interpretation of the score. What kinds of variability (variance) are there in this?

Because the SVT/PVT measures involve tasks that are very easy to perform or target unusual or infrequent symptom endorsement, the scores do not follow a normal distribution. The PVT tasks are not graded with some kind of escalating scale of difficulty and for those that have a "hard" or challenging task it is not incrementally scaled. So none of these scores follows a normal distribution. This is why SVT/PVT proponents like to argue the "all-or-none" cut-score approach to defining pass and fail. So for those who pass there is almost no variance in the scores.

As explained in the Bigler (2012) review there is a bimodal distribution of those who "failed" in the study by Locke and colleagues (2008) study. One peak is clearly in the "grey

zone" close to a pass and the other peak is closer to scores trending toward the chance range. But even in the Locke study only three patients scored at chance level, with none performing substantially below chance. The further away from the cut-point, the greater the suspicion that the test results may not credibly represent true ability levels. However, there are enough circumstances on an individual basis where this could be interpretatively ambiguous. Once scores are at or below chance, then non-credible performance is most probable to definite unless a well-commented intellectual disability or case of dementia is present.

How should one think about "near misses"?

In my clinical experience, near misses should not be interpreted as invalid responding unless some clear-cut evidence indicates otherwise. For most of the major disorders, as explained above, there is ample evidence to allow the clinician to NOT interpret "near misses" as unequivocal indicators of invalid performance.

For scores that are above chance but substantially lower than the cut score, it is considerably more challenging to explain credible performance.

When the test results indicate poor effort, what other methods or information can be used to validate/invalidate test results?

There are internal, embedded measures that can be checked along with the external SVT/PVT measures to validate overall whether there is credible performance in some areas. As already stated the embedded SVT/PVT measures are not available for all domains assessed so only a limited cross checking between external and internal validity measures can be made. In some cases, SVT/PVT failure may be specific to just one domain, like memory, with other areas of cognitive functioning performed credibly. Unfortunately, much of this approach relies solely on clinical judgment.

As explained above, in the study by Suchy and colleagues, the examinee may be confronted if a SVT/PVT measure has been failed. In the Comprehensive Clinic at Brigham Young University (BYU), a community based full-service psychological/neuropsychological assessment service, typically a minimum of two SVT/PVT measures are administered for a comprehensive neuropsychological examination. One SVT/PVT measure will be administered early in the assessment. If we find an examinee exhibiting poor engagement in the testing as

indicated by below cut-score SVT/PVT performance, we will confront the individual in a clinically meaningful way to encourage greater test engagement and performance. What is communicated to the patient depends partially on how substantial the poor performance is on the SVT/PVT measure. If the examinee passes subsequent SVT/PVT measures, we will continue with the assessment. If SVT/PVT failure continues and if we have no other explanation (e.g., schizophrenia, dementia) for the failed performance, then we may discontinue testing.

The "mental status" section of the BYU Comprehensive Clinic report contains a specific section on test validity wherein an elaborate description of both external and embedded measures will be outlined and interpreted.

When the test results indicate poor effort, but there is also evidence of significant functional limitation, how should a disability examiner interpret these results?

As indicated in this review, clear-cut functional limitation (i.e., hemiplegia) especially associated with positive objective neuroimaging findings trump SVT/PVT findings. In cases that meet SSA criteria for "Compassionate Allowances," it would seem that SVT/PVT measures add little to the evaluation process, except if there was suspicion of malingering.

How should the disability examiner rule out alternative explanations for the results rendered by the technique?

This review provides perspective on how alternative explanations should be explored. As mentioned in this review, the tendency has been to not explore alternative explanations and to simply not interpret the findings. Some of this will require better training on the potential full spectrum of SVT/PVT interpretation than what has been the tradition.

Are SVTs sensitive to or do they permit a determination of exaggeration as opposed to malingering?

If, in a National Library of Medicine (pubmed.gov) search, one pairs the term "MMPI" , as the SVT measure, with "exaggeration," 51 publications appear; however if "MMPI" is paired with "malingering," 227 publications are listed. From this rather unscientific comparison, more psychologists would use the malingering classification than exaggeration. However, as already discussed, once a neuropsychiatric disorder moves into the spectrum of somatic symptom

disorder and associated sickness behavior (Maes et al., 2012), differentiating and defining what may be exaggeration versus the individuals heightened reporting of symptoms is really unknown.

If specific effort testing is not done, are there other indicators during the examination that indicate poor effort?

This is the reason for new test development with embedded measures so that each measure administered has some element that will permit some statement about validity of performance. Abundant research shows that although clinical judgment is important (Bigler, 1990), it is not sufficient as a singular method to diagnose insufficient effort, suboptimal performance, or malingering in most situations (Faust and Guilmette, 1990).

What does one do with clinical presentations that are not "typical"?

Lots of controversy in the SVT/PVT literature surrounds concussion and mild TBI research. Twenty years ago, Stuss (1995) offered several points that apply not only to understanding mild TBI and neurocognitive and neurobehavioral sequelae, but also how to address SVT/PVT findings, their interpretation, and controversies. He outlined five principles paraphrased as follows: (1) Everything must make sense, (2) the severity of the injury must be defined not by the number of symptoms endorsed but by the acute characteristics of the injury, (3) expected outcomes are defined by well-designed studies that provide appropriate baseline information (4) conclusions should be logical, explain the findings and resonate with known outcomes for the injury/disorder and (5) the examiner has a clinical responsibility for ethical interpretation.

Twenty-first century medical knowledge and psychological sophistication should provide ample information on how to address atypical cases, but unfortunately to often clinicians have merely dismissed the case where failed SVT/PVT scores have occurred rather than explored how to manage the atypical case.

FINAL CONCLUSIONS

Psychological and neuropsychological test developers understand that embedded validity measures need to be a part of the next generation of test development. Attempting to retrofit old, established psychological and neuropsychological methods with external SVT/PVT tasks will likely never achieve a satisfactory endpoint. In my opinion this is a dead-end approach and

should be abandoned. From the neuropsychological perspective, Larrabee (2014b) provides what I would consider exactly the correct focus for future assessment technology in what he refers to as an Ability-Focused Neuropsychological Test Battery (AFB). Importantly Larrabee emphasizes the need to standardize the AFB to criterion groups such as Alzheimer's disease, stroke, TBI, and the like, such that the measures have appropriate levels of sensitivity in detecting neurocognitive impairments associated with each disorder as well as what may be unique patterns of neurocognitive deficits within each disorder. Mostly importantly, each domain measure of cognitive functioning would be developed "… based on various criteria for validity including sensitivity to presence of disorder, sensitivity to severity of disorder, correlation with important activities of daily living, and containing embedded/derived measures of performance validity" (p. e-pub). In other words, either within individual tests or within a domain there will be internal validity metrics, as there should be.

There is no such battery in existence at this time. However, NIH has recognized the need to more uniformly assess neurocognitive outcome for a variety of conditions that require clinical monitoring where cognition is the outcome variable. As a means to achieve this goal, NIH has funded the development of the "NIH Toolbox for Assessment of Neurological and Behavioral Function" (Gershon et al., 2013) with the goal " to develop a set of state-of-the-art measurement tools to enhance collection of data in large cohort studies and to advance the biomedical research enterprise" (p. S2, Gershon et al. 2013). In part this toolbox is computer based and hopefully will eventually have built-in SVT/PVT measures. With the toolbox being in the public domain and with the potential for web-based modifications of databases, it has the potential for generating the largest normative dataset known to the assessment literature. This would permit directing assessment methods and findings, including validity performance to specific conditions and developing unique algorithms for each condition.

In clinical outcome research involving schizophrenia, there has been the development of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (MCCB) (MCCB, Green et al., 2004). This battery combined with a computer-based virtual assessment method to assess functional capacity appears to have considerable potential in defining the cognitive, neurobehavioral, emotional, and motivational deficits in schizophrenia (Ruse et al., 2014). The next step would be to computerize the entire

assessment procedure with embedded SVT/PVT measures and standardize it across neuropsychiatric domains.

Given the number of disability evaluations that the SSA processes annually, something along these lines is needed for uniformity and interpretability. At this point in time, I anticipate that psychologists performing assessments to be used by SSA will include some form of a SVT/PVT, but as this review demonstrates there could be no specific recommendation as to which specific measures should be given.

The SSA should also be apprised that there is a burgeoning literature on electrophysiological, eye-tracking, and neuroimaging approaches to detect malingering. None of this literature was reviewed. In the future neurocognitive assessments integrated with these approaches will likely address all of the current limitations of SVT/PVT methods elaborated in this review.

REFERENCES

Alfano, D. P. 2006. Emotional and pain-related factors in neuropsychological assessment following mild traumatic brain injury. *Brain and cognition* 60(2):194-196.

Allen, M. D., T. C. Wu, and E. D. Bigler. 2011. Traumatic brain injury alters word memory test performance by slowing response time and increasing cortical activation: An fmri study of a symptom validity test. *Psychological Injury and Law* 4:140-146.

Anastasi, A. 1954. *Psychological testing*. New York: Macmillan.

Avery, R., M. Startup, and K. Calabria. 2009. The role of effort, cognitive expectancy appraisals and coping style in the maintenance of the negative symptoms of schizophrenia. *Psychiatry research* 167(1-2):36-46.

Barhon, L. I., J. Batchelor, S. Meares, E. Chekaluk, and E. A. Shores. 2014. A comparison of the degree of effort involved in the tomm and the acs word choice test using a dual-task paradigm. *Applied neuropsychology. Adult*:1-10.

Barr, W. B. 2013. Differential diagnosis of psychological factors evoked by pain presentations. *The Clinical neuropsychologist* 27(1):17-29.

Bashem, J. R., L. J. Rapport, J. B. Miller, R. A. Hanks, B. N. Axelrod, and S. R. Millis. 2014. Comparisons of five performance validity indices in bona fide and simulated traumatic brain injury. *The Clinical neuropsychologist* 28(5):851-875.

Bass, C., and P. Halligan. 2014. Factitious disorders and malingering: Challenges for clinical assessment and management. *Lancet* 383(9926):1422-1432.

Batt, K., E. A. Shores, and E. Chekaluk. 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 : JINS* 14(6):1074-1080.

Bauer, L., and R. J. McCaffrey. 2006. Coverage of the test of memory malingering, victoria symptom validity test, and word memory test on the internet: Is test security threatened?

Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists 21(1):121-126.
Bellack, A. S., K. T. Mueser, R. L. Morrison, A. Tierney, and K. Podell. 1990. Remediation of cognitive deficits in schizophrenia. *Am J Psychiatry* 147(12):1650-1655.
Bernard, L. C. 1990. Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of clinical and experimental neuropsychology* 12(5):715-728.
Bernard, L. C., and W. Fowler. 1990. Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on rey's task to detect malingering. *Journal of clinical psychology* 46(4):432-436.
Berry, D. T., J. J. Adams, C. D. Clark, S. R. Thacker, T. L. Burger, M. W. Wetter, and R. A. Baer. 1996. Detection of a cry for help on the mmpi-2: An analog investigation. *Journal of personality assessment* 67(1):26-36.
Berthelson, L., S. S. Mulchan, A. P. Odland, L. J. Miller, and W. Mittenberg. 2013. False positive diagnosis of malingering due to the use of multiple effort tests. *Brain injury : [BI]* 27(7-8):909-916.
Bianchini, K. J., K. W. Greve, and J. M. Love. 2003. Definite malingered neurocognitive dysfunction in moderate/severe traumatic brain injury. *The Clinical neuropsychologist* 17(4):574-580.
Bianchini, K. J., C. W. Mathias, and K. W. Greve. 2001. Symptom validity testing: A critical review. *The Clinical neuropsychologist* 15(1):19-45.
Bigler, E. D. 1990. Neuropsychology and malingering: Comment on faust, hart, and guilmette (1988). *Journal of consulting and clinical psychology* 58(2):244-247.
———. 2006. Can author bias be determined in forensic neuropsychology research published in archives of clinical neuropsychology? *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 21(5):503-508.
———. 2012. Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society : JINS* 18(4):632-640.
———. 2014. Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain injury : [BI]*:1-16.
Bigler, E. D., and M. Brooks. 2009. Traumatic brain injury and forensic neuropsychology. *The Journal of head trauma rehabilitation* 24(2):76-87.
Bigler, E. D., R. R. Green, T. J. Farrer, J. C. Roper, and J. B. Millward. 2009. The rigor of research design and "forensic" publications in neuropsychological research. *Psychological Injury and Law* 2:43-52.
Bijleveld, E., R. Custers, S. Van der Stigchel, H. Aarts, P. Pas, and M. Vink. 2014. Distinct neural responses to conscious versus unconscious monetary reward cues. *Human brain mapping*.
Bilder, R. M., C. A. Sugar, and G. S. Hellemann. 2015. Cumulative false positive rates given multiple performance validity tests:
Commentary on davis and millis (2014) and larrabee (2014). *The Clinical neuropsychologist* in press:in press.
Binder, L. M., and L. Pankratz. 1987. Neuropsychological evidence of a factitious memory complaint. *Journal of clinical and experimental neuropsychology* 9(2):167-171.

Blank, J., L. Evered, D. Watson, and R. Ruff. 2014. C-87malingering madness: Distress as a diagnostic alternative. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):605.

Bolinger, E., C. Reese, J. Suhr, and G. J. Larrabee. 2014. Susceptibility of the mmpi-2-rf neurological complaints and cognitive complaints scales to over-reporting in simulated head injury. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(1):7-15.

Bortnik, K. E., M. D. Horner, and D. L. Bachman. 2013. Performance on standard indexes of effort among patients with dementia. *Applied neuropsychology. Adult*.

Bowden, S. C., E. A. Shores, and J. L. Mathias. 2006. Does effort suppress cognition after traumatic brain injury? A re-examination of the evidence for the word memory test. *The Clinical neuropsychologist* 20(4):858-872.

Burton, V., R. Vilar-Lopez, and A. E. Puente. 2012. Measuring effort in neuropsychological evaluations of forensic cases of spanish speakers. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 27(3):262-267.

Bush, S. S., R. M. Ruff, A. I. Troster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity nan policy & planning committee. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 20(4):419-426.

Butcher, J. N. 2010. Personality assessment from the nineteenth to the early twenty-first century: Past achievements and contemporary challenges. *Annual review of clinical psychology* 6:1-20.

Butcher, J. N., P. A. Arbisi, M. M. Atlis, and J. L. McNulty. 2003. The construct validity of the lees-haley fake bad scale. Does this scale measure somatic malingering and feigned emotional distress? *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 18(5):473-485.

Cahn, D. A., E. V. Sullivan, P. K. Shear, L. Marsh, R. Fama, K. O. Lim, J. A. Yesavage, J. R. Tinklenberg, and A. Pfefferbaum. 1998. Structural mri correlates of recognition memory in alzheimer's disease. *Journal of the International Neuropsychological Society : JINS* 4(2):106-114.

Chafetz, M. 2011. Reducing the probability of false positives in malingering detection of social security disability claimants. *The Clinical neuropsychologist* 25(7):1239-1252.

Chafetz, M., and J. Underhill. 2013. Estimated costs of malingered disability. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 28(7):633-639.

Chafetz, M. D. 2012. The a-test: A symptom validity indicator embedded within a mental status examination for social security disability. *Applied neuropsychology. Adult* 19(2):121-126.

Chafetz, M. D., and A. Biondolillo. 2012. Validity issues in atkins death cases. *The Clinical neuropsychologist* 26(8):1358-1376.

Chafetz, M. D., E. Prentkowski, and A. Rao. 2011. To work or not to work: Motivation (not low iq) determines symptom validity test findings. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 26(4):306-313.

Clark, A. L., M. M. Amick, C. Fortier, W. P. Milberg, and R. E. McGlinchey. 2014. Poor performance validity predicts clinical characteristics and cognitive test performance of oef/oif/ond veterans in a research setting. *The Clinical neuropsychologist* 28(5):802-825.

Cottingham, M. E., T. L. Victor, K. B. Boone, E. A. Ziegler, and M. Zeller. 2014. Apparent effect of type of compensation seeking (disability versus litigation) on performance validity test scores may be due to other factors. *The Clinical neuropsychologist* 28(6):1030-1047.

Darwin, C. 1872. *The expression of the emotions in man and animals*. London: John Murray.

Davis, J. J., T. S. McHugh, A. D. Bagley, B. N. Axelrod, and R. A. Hanks. 2011. Cross-validation of picture completion effort indices in personal injury litigants and disability claimants. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 26(8):768-773.

Dean, A. C., T. L. Victor, K. B. Boone, and G. Arnold. 2008. The relationship of iq to effort test performance. *The Clinical neuropsychologist* 22(4):705-722.

Dean, A. C., T. L. Victor, K. B. Boone, L. M. Philpott, and R. A. Hess. 2009. Dementia and effort test performance. *The Clinical neuropsychologist* 23(1):133-152.

DenBoer, J. W., and S. Hall. 2007. Neuropsychological test performance of successful brain injury simulators. *The Clinical neuropsychologist* 21(6):943-955.

Dere, J., J. Sun, Y. Zhao, T. J. Persson, X. Zhu, S. Yao, R. M. Bagby, and A. G. Ryder. 2013. Beyond "somatization" and "psychologization": Symptom-level variation in depressed han chinese and euro-canadian outpatients. *Frontiers in psychology* 4:377.

Deright, J., and D. A. Carone. 2013. Assessment of effort in children: A systematic review. *Child neuropsychology : a journal on normal and abnormal development in childhood and adolescence*.

Dodrill, C. B. 2008. Do patients with psychogenic nonepileptic seizures produce trustworthy findings on neuropsychological tests? *Epilepsia* 49(4):691-695.

Drane, D. L., D. J. Williamson, E. S. Stroup, M. D. Holmes, M. Jung, E. Koerner, N. Chaytor, A. J. Wilensky, and J. W. Miller. 2006. Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia* 47(11):1879-1886.

Duquin, J. A., B. A. Parmenter, and R. H. Benedict. 2008. Influence of recruitment and participation bias in neuropsychological research among ms patients. *Journal of the International Neuropsychological Society : JINS* 14(3):494-498.

Etherton, J. L., K. J. Bianchini, K. W. Greve, and M. A. Ciota. 2005. Test of memory malingering performance is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 20(3):375-384.

Faust, D., and T. J. Guilmette. 1990. To say it's not so doesn't prove that it isn't: Research on the detection of malingering. Reply to bigler. *Journal of consulting and clinical psychology* 58(2):248-250.

Fervaha, G., K. K. Zakzanis, G. Foussias, A. Graff-Guerrero, O. Agid, and G. Remington. 2014. Motivational deficits and cognitive test performance in schizophrenia. *JAMA psychiatry* 71(9):1058-1065.

Ford, E. B. 2006. Lie detection: Historical, neuropsychiatric and legal dimensions. *International journal of law and psychiatry* 29(3):159-177.

Fox, D. D. 2011. Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical neuropsychologist* 25(3):488-495.

Franzen, M. D., and N. Martin. 1996. Do people with knowledge fake better? *Applied neuropsychology* 3(2):82-85.

Frazier, T. W., E. A. Youngstrom, R. I. Naugle, K. A. Haggerty, and R. M. Busch. 2007. The latent structure of cognitive symptom exaggeration on the victoria symptom validity test. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 22(2):197-211.

Gass, C. S., and A. P. Odland. 2012. Minnesota multiphasic personality inventory-2 revised form symptom validity scale-revised (mmpi-2-rf fbs-r; also known as fake bad scale): Psychometric characteristics in a nonlitigation neuropsychological setting. *Journal of clinical and experimental neuropsychology* 34(6):561-570.

———. 2014. Mmpi-2 symptom validity (fbs) scale: Psychometric characteristics and limitations in a veterans affairs neuropsychological setting. *Applied neuropsychology. Adult* 21(1):1-8.

Gershon, R. C., M. V. Wagster, H. C. Hendrie, N. A. Fox, K. F. Cook, and C. J. Nowinski. 2013. Nih toolbox for assessment of neurological and behavioral function. *Neurology* 80(11 Suppl 3):S2-6.

Gervais, R. O., Y. S. Ben-Porath, D. B. Wygant, and P. Green. 2007. Development and validation of a response bias scale (rbs) for the mmpi-2. *Assessment* 14(2):196-208.

Gervais, R. O., Y. S. Ben-Porath, D. B. Wygant, and M. Sellbom. 2010. Incremental validity of the mmpi-2-rf over-reporting scales and rbs in assessing the veracity of memory complaints. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 25(4):274-284.

Glanzer, M., J. K. Adams, G. J. Iverson, and K. Kim. 1993. The regularities of recognition memory. *Psychological review* 100(3):546-567.

Goldberg, J. O., and H. R. Miller. 1986. Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *Journal of clinical psychology* 42(5):792-795.

Gorissen, M., J. C. Sanz, and B. Schmand. 2005. Effort and cognition in schizophrenia patients. *Schizophrenia research* 78(2-3):199-208.

Goslin, D. A. 1968. Standardized ability tests and testing. Major issues and the validity of current criticisms of tests are discussed. *Science* 159(3817):851-855.

Gough, H. G. 1950. The f minus k dissimulation index for the minnesota multiphasic personality inventory. *Journal of consulting psychology* 14(5):408-413.

Graham, J. R., D. Watts, and R. E. Timbrook. 1991. Detecting fake-good and fake-bad mmpi-2 profiles. *Journal of personality assessment* 57(2):264-277.

Green, M. F., S. Ganzell, P. Satz, and J. F. Vaclav. 1990. Teaching the wisconsin card sorting test to schizophrenic patients. *Archives of general psychiatry* 47(1):91-92.

Green, M. F., K. H. Nuechterlein, J. M. Gold, D. M. Barch, J. Cohen, S. Essock, W. S. Fenton, F. Frese, T. E. Goldberg, R. K. Heaton, R. S. Keefe, R. S. Kern, H. Kraemer, E. Stover, D. R. Weinberger, S. Zalcman, and S. R. Marder. 2004. Approaching a consensus cognitive battery for clinical trials in schizophrenia: The nimh-matrics conference to select cognitive domains and test criteria. *Biological psychiatry* 56(5):301-307.

Greiffenstein, M., R. Gervais, W. J. Baker, L. Artiola, and H. Smith. 2013. Symptom validity testing in medically unexplained pain: A chronic regional pain syndrome type 1 case series. *The Clinical neuropsychologist* 27(1):138-147.

Greiffenstein, M. F., W. J. Baker, B. Axelrod, E. A. Peck, and R. Gervais. 2004. The fake bad scale and mmpi-2 f-family in detection of implausible psychological trauma claims. *The Clinical neuropsychologist* 18(4):573-590.

Griffin, G. A., J. Normington, R. May, and D. Glassmire. 1996. Assessing dissimulation among social security disability income claimants. *Journal of consulting and clinical psychology* 64(6):1425-1430.

Guilmette, T. J., W. Whelihan, F. R. Sparadeo, and G. Buongiorno. 1994. Validity of neuropsychological test results in disability evaluations. *Perceptual and motor skills* 78(3 Pt 2):1179-1186.

Guilmette, T. J., W. M. Whelihan, K. J. Hart, F. R. Sparadeo, and G. Buongiorno. 1996. Order effects in the administration of a forced-choice procedure for detection of malingering in disability claimants' evaluations. *Perceptual and motor skills* 83(3 Pt 1):1007-1016.

Hampson, N. E., S. Kemp, A. K. Coughlan, C. J. Moulin, and B. B. Bhakta. 2014. Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. *Applied neuropsychology. Adult* 21(3):183-194.

Harrison, A. G., and I. Armstrong. 2014. Wisc-iv unusual digit span performance in a sample of adolescents with learning disabilities. *Applied neuropsychology. Child* 3(2):152-160.

Harvey, P. D., R. K. Heaton, W. T. Carpenter, Jr., M. F. Green, J. M. Gold, and M. Schoenbaum. 2012. Functional impairment in people with schizophrenia: Focus on employability and eligibility for disability compensation. *Schizophrenia research* 140(1-3):1-8.

Hase, H. D., and L. R. Goldberg. 1967. Comparative validity of different strategies of constructing personality inventory scales. *Psychological bulletin* 67(4):231-248.

Hathaway, S. R. 1947. A coding system for mmpi profile classification. *Journal of consulting psychology* 11(6):334-337.

Hathaway, S. R., and P. F. Briggs. 1957. Some normative data on new mmpi scales. *Journal of clinical psychology* 13(4):364-368.

Hathaway, S. R., and J. C. McKinley. 1940. A multiphasic personality schedule(minnesota): I. Construction of the schedule. *Journal of Psychology* 10:249-254.

Hathaway, S. R., and J. C. McKinley. 1943. *Manual for the minnesota multiphasic personality inventory*. Minneapolis: University of Minnesota.

Hays, J. R., J. Emmons, and K. A. Lawson. 1993. Psychiatric norms for the rey 15-item visual memory test. *Perceptual and motor skills* 76(3 Pt 2):1331-1334.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, and S. R. Millis. 2009. American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical neuropsychologist* 23(7):1093-1129.

Horwitz, J. E., and R. J. McCaffrey. 2006. A review of internet sites regarding independent medical examinations: Implications for clinical neuropsychological practitioners. *Applied neuropsychology* 13(3):175-179.

Hunt, S., J. C. Root, and B. L. Bascetta. 2014. Effort testing in schizophrenia and schizoaffective disorder: Validity indicator profile and test of memory malingering performance characteristics. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(2):164-172.

Hyer, L., M. Woods, W. R. Harrison, P. Boudewyns, and W. C. O'Leary. 1989. Mmpi f-k index among hospitalized vietnam veterans. *Journal of clinical psychology* 45(2):250-254.

Johnson-Greene, D., L. Brooks, and T. Ference. 2013. Relationship between performance validity testing, disability status, and somatic complaints in patients with fibromyalgia. *The Clinical neuropsychologist* 27(1):148-158.

Johnson, J. L., C. G. Bellah, T. Dodge, W. Kelley, and M. M. Livingston. 1998. Effect of warning on feigned malingering on the wais-r in college samples. *Perceptual and motor skills* 87(1):152-154.

Johnson, J. L., and K. Lesniak-Karpiak. 1997. The effect of warning on malingering on memory and motor tasks in college samples. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 12(3):231-238.

Jones, A., and M. V. Ingram. 2011. A comparison of selected mmpi-2 and mmpi-2-rf validity scales in assessing effort on cognitive tests in a military sample. *The Clinical neuropsychologist* 25(7):1207-1227.

Jones, A., M. V. Ingram, and Y. S. Ben-Porath. 2012. Scores on the mmpi-2-rf scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical neuropsychologist* 26(5):790-815.

Jung, J. 1967. Cued versus non-cued incidental recall of successive word associations. *Canadian journal of psychology* 21(3):196-203.

Kim, M. S., K. B. Boone, T. Victor, S. D. Marion, S. Amano, M. E. Cottingham, E. A. Ziegler, and M. A. Zeller. 2010. The warrington recognition memory test for words as a measure of response bias: Total score and response time cutoffs developed on "real world" credible and noncredible subjects. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 25(1):60-70.

King, P. M. 1998. Analysis of approaches to detection of sincerity of effort through grip strength measurement. *Work* 10(1):9-13.

Lafargue, G., and N. Franck. 2009. Effort awareness and sense of volition in schizophrenia. *Consciousness and cognition* 18(1):277-289.

Lange, R. T., T. Brickell, L. M. French, B. Ivins, A. Bhagwat, S. Pancholi, and G. L. Iverson. 2013. Risk factors for postconcussion symptom reporting after traumatic brain injury in u.S. Military service members. *Journal of neurotrauma* 30(4):237-246.

Larrabee, G. J. 1997. Neuropsychological outcome, post concussion symptoms, and forensic considerations in mild closed head trauma. *Seminars in clinical neuropsychiatry* 2(3):196-206.

———. 2003a. Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical neuropsychologist* 17(3):410-425.

———. 2003b. Exaggerated mmpi-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 18(6):673-686.

———. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society : JINS* 18(4):625-630.

———. 2014a. False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(4):364-373.

———. 2014b. Test validity and performance validity: Considerations in providing a framework for development of an ability-focused neuropsychological test battery. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*.

Larson, E., F. Zollman, B. Kondiles, and C. Starr. 2013. Memory deficits, postconcussive complaints, and posttraumatic stress disorder in a volunteer sample of veterans. *Rehabilitation psychology* 58(3):245-252.

Leavitt, F. 1985. The value of the mmpi conversion 'v' in the assessment of psychogenic pain. *Journal of psychosomatic research* 29(2):125-131.
Lees-Haley, P. R., L. T. English, and W. J. Glenn. 1991. A fake bad scale on the mmpi-2 for personal injury claimants. *Psychological reports* 68(1):203-210.
Lewis, D. O., C. A. Yeager, P. Blake, B. Bard, and M. Strenziok. 2004. Ethics questions raised by the neuropsychiatric, neuropsychological, educational, developmental, and family characteristics of 18 juveniles awaiting execution in texas. *The journal of the American Academy of Psychiatry and the Law* 32(4):408-429.
Lezak, M. D., D. B. Howieson, E. D. Bigler, and D. Tranel. 2012. *Neuropsychological assessment*. 5th ed. New York: Oxford University Press.
Locke, D. E., J. S. Smigielski, M. R. Powell, and S. R. Stevens. 2008. Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation* 23(3):273-281.
Loring, D. W., G. P. Lee, and K. J. Meador. 2005. Victoria symptom validity test performance in non-litigating epilepsy surgery candidates. *Journal of clinical and experimental neuropsychology* 27(5):610-617.
Loring, D. W., S. E. Marino, D. L. Drane, D. Parfitt, G. R. Finney, and K. J. Meador. 2011. Lorazepam effects on word memory test performance: A randomized, double-blind, placebo-controlled, crossover trial. *The Clinical neuropsychologist* 25(5):799-811.
Maes, M., M. Berk, L. Goehler, C. Song, G. Anderson, P. Galecki, and B. Leonard. 2012. Depression and sickness behavior are janus-faced responses to shared inflammatory pathways. *BMC medicine* 10:66.
Manly, J. J. 2008. Critical issues in cultural neuropsychology: Profit from diversity. *Neuropsychology review* 18(3):179-183.
Marcopulos, B. A., B. A. Caillouet, C. M. Bailey, C. Tussey, J. A. Kent, and R. Frederick. 2014. Clinical decision making in response to performance validity test failure in a psychiatric setting. *The Clinical neuropsychologist* 28(4):633-652.
McCarthy, R. A., M. D. Kopelman, and E. K. Warrington. 2005. Remembering and forgetting of semantic knowledge in amnesia: A 16-year follow-up investigation of rfr. *Neuropsychologia* 43(3):356-372.
McGrath, R. E., M. Mitchell, B. H. Kim, and L. Hough. 2010. Evidence for response bias as a source of error variance in applied assessment. *Psychological bulletin* 136(3):450-470.
Merydith, S. P., and F. H. Wallbrown. 1991. Reconsidering response sets, test-taking attitudes, dissimulation, self-deception, and social desirability. *Psychological reports* 69(3 Pt 1):891-905.
Meyers, J. E., M. Volbrecht, B. N. Axelrod, and L. Reinsch-Boothby. 2011. Embedded symptom validity tests and overall neuropsychological test performance. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 26(1):8-15.
Michaels, D. 2008. *Doubt is their product*. New York: Oxford University Press.
Miele, A. S., J. H. Gunner, J. K. Lynch, and R. J. McCaffrey. 2012. Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 27(1):10-22.
Milgram, S. 1963. Behavioral study of obedience. *Journal of abnormal psychology* 67:371-378.

Millis, S. R. 2009. Methodological challenges in assessment of cognition following mild head injury: Response to malojcic et al. 2008. *Journal of neurotrauma* 26(12):2409-2410.
Miro, J., R. Nieto, and A. Huguet. 2008. The catalan version of the pain catastrophizing scale: A useful instrument to assess catastrophic thinking in whiplash patients. *The journal of pain : official journal of the American Pain Society* 9(5):397-406.
Mooney, G., J. Speed, and S. Sheppard. 2005. Factors related to recovery after mild traumatic brain injury. *Brain injury : [BI]* 19(12):975-987.
Morel, K. R., and K. C. Marshman. 2008. Critiquing symptom validity tests for posttraumatic stress disorder: A modification of hartman's criteria. *Journal of anxiety disorders* 22(8):1542-1550.
Morel, K. R., and B. E. Shepherd. 2008. Developing a symptom validity test for posttraumatic stress disorder: Application of the binomial distribution. *Journal of anxiety disorders* 22(8):1297-1302.
Morra, L., J. Gold, K. Ossenfort, and G. Strauss. 2014. C-78predicting insufficient effort in schizophrenia using the repeated battery for the assessment of neuropsychological status effort index. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):602-603.
Morris, R. G., S. Abrahams, and C. E. Polkey. 1995. Recognition memory for words and faces following unilateral temporal lobectomy. *The British journal of clinical psychology / the British Psychological Society* 34 (Pt 4):571-576.
Mosti, C., B. Locklair, A. Lyon, and T. Swirsky-Sacchetti. 2014. C-47sleep disturbance and symptom validity testing in mild traumatic brain injury patients. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):591.
Murrie, D. C., M. T. Boccaccini, L. A. Guarnera, and K. A. Rufino. 2013. Are forensic experts biased by the side that retained them? *Psychological science* 24(10):1889-1897.
Nave, P., S. Mooney, J. Raintree, E. Seats, E. Hamrick, K. Jihad, and E. Taylor. 2014. C-82military medical evaluation board impact on credibility of cognitive and psychiatric results. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):604.
Nelson, N. W., K. Boone, A. Dueck, L. Wagener, P. Lu, and C. Grills. 2003. Relationships between eight measures of suspect effort. *The Clinical neuropsychologist* 17(2):263-272.
Nelson, N. W., J. J. Sweet, and R. L. Heilbronner. 2007. Examination of the new mmpi-2 response bias scale (gervais): Relationship with mmpi-2 validity scales. *Journal of clinical and experimental neuropsychology* 29(1):67-72.
Neurology, A. A. o. 2011. *Clinical practice guideline process manual*. St. Paul, MN: The American Academy of Neurology.
Nisbet, H., R. Siegert, M. Hunt, and N. Fairley. 1996. Improving schizophrenic in-patients' wisconsin card-sorting performance. *The British journal of clinical psychology / the British Psychological Society* 35 (Pt 4):631-633.
Nussbaum, K., A. M. Schneidmuhl, and J. W. Shaffer. 1969a. Psychiatric assessment in the social security program of disability insurance. *Am J Psychiatry* 126(6):897-899.
Nussbaum, K., J. W. Shaffer, and A. M. Schneidmuhl. 1969b. Psychological assessment in the social security program of disability insurance. *Am Psychol* 24(9):869-872.

Paganini-Hill, A., B. Ducey, and M. Hawk. 2013. Responders versus nonresponders in a dementia study of the oldest old: The 90+ study. *American journal of epidemiology* 177(12):1452-1458.

Pankratz, L. 1979. Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of consulting and clinical psychology* 47(2):409-410.

Pankratz, L., L. M. Binder, and L. M. Wilcox. 1987. Evaluation of an exaggerated somatosensory deficit with symptom validity testing. *Archives of neurology* 44(8):798.

Pankratz, L., A. Fausti, and S. Peed. 1975. A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of consulting and clinical psychology* 43(3):421-422.

Pedersen, H., L. Schwent Shultz, B. Roper, G. Crucian, and E. Crouse. 2014. C-80is deceptive language necessary to detect posttraumatic stress disorder (ptsd) symptom exaggeration? A look at the morel emotional numbing test (ment). *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):603.

Pessoa, L., and J. B. Engelmann. 2010. Embedding reward signals into perception and cognition. *Frontiers in neuroscience* 4.

Peterson, G. W., D. A. Clark, and B. Bennett. 1989. The utility of mmpi subtle, obvious scales for detecting fake good and fake bad response sets. *Journal of clinical psychology* 45(4):575-583.

Poole, C. J. 2010. Illness deception and work: Incidence, manifestations and detection. *Occupational medicine* 60(2):127-132.

Rey, A. 1958. *L'examen clinique en psychologie [the psychological examination]*. Paris: Presses Universitaires de France.

Reznek, L. 2005. The rey 15-item memory test for malingering: A meta-analysis. *Brain injury : [BI]* 19(7):539-543.

Riba, J., A. Rodriguez-Fornells, T. F. Munte, and M. J. Barbanoj. 2005. A neurophysiological study of the detrimental effects of alprazolam on human action monitoring. *Brain research. Cognitive brain research* 25(2):554-565.

Rivera Mindt, M., D. Byrd, P. Saez, and J. Manly. 2010. Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. *The Clinical neuropsychologist* 24(3):429-453.

Rose, F. E., S. Hall, A. D. Szalda-Petree, and P. J. Bach. 1998. A comparison of four tests of malingering and the effects of coaching. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 13(4):349-363.

Ruse, S. A., P. D. Harvey, V. G. Davis, A. S. Atkins, K. H. Fox, and R. S. Keefe. 2014. Virtual reality functional capacity assessment in schizophrenia: Preliminary data regarding feasibility and correlations with cognitive and functional capacity performance. *Schizophrenia research. Cognition* 1(1):e21-e26.

Samet, J. M., and T. A. Burke. 2001. Turning science into junk: The tobacco industry and passive smoking. *American journal of public health* 91(11):1742-1744.

Schenk, K., and K. A. Sullivan. 2010. Do warnings deter rather than produce more sophisticated malingering? *Journal of clinical and experimental neuropsychology* 32(7):752-762.

Scotland, J. L., R. Al-Shahi Salman, I. J. Deary, and I. R. Whittle. 2009. Recruitment difficulties in brain tumour patients cause participation bias: Findings from a neuropsychological

study of adult inpatients with supratentorial intracranial tumours. *Acta neurochirurgica* 151(10):1191-1195.
Sellbom, M., D. Wygant, and M. Bagby. 2012. Utility of the mmpi-2-rf in detecting non-credible somatic complaints. *Psychiatry research* 197(3):295-301.
Silver, J. M. 2012. Effort, exaggeration and malingering after concussion. *Journal of neurology, neurosurgery, and psychiatry* 83(8):836-841.
Simmons, R. D. 2010. Life issues in multiple sclerosis. *Nature reviews. Neurology* 6(11):603-610.
Simonsen, J. C. 1996. Validation of sincerity of effort. *Journal of back and musculoskeletal rehabilitation* 6(3):289-295.
Simpson, J. R., ed. 2012. *Neuroimaging in forensic psychiatry: From the clinic to the courtroom*. New York: Wiley-Blackwell.
Slick, D. J., E. M. Sherman, and G. L. Iverson. 1999. Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical neuropsychologist* 13(4):545-561.
Slick, D. J., J. E. Tan, E. H. Strauss, and D. F. Hultsch. 2004. Detecting malingering: A survey of experts' practices. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 19(4):465-473.
Smith, G. A., R. C. Nelson, S. J. Sadoff, and A. M. Sadoff. 1989. Assessing sincerity of effort in maximal grip strength tests. *American journal of physical medicine & rehabilitation / Association of Academic Physiatrists* 68(2):73-80.
Smith, K., K. Boone, T. Victor, D. Miora, M. Cottingham, E. Ziegler, M. Zeller, and M. Wright. 2014. Comparison of credible patients of very low intelligence and non-credible patients on neurocognitive performance validity indicators. *The Clinical neuropsychologist* 28(6):1048-1070.
Sollman, M. J., and D. T. Berry. 2011. Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 26(8):774-789.
Standing, L. 1973. Learning 10,000 pictures. *The Quarterly journal of experimental psychology* 25(2):207-222.
Stevens, A., K. Schneider, B. Liske, L. Hermle, H. Huber, and G. Hetzel. 2014. Is suboptimal cogntive performance in schizophrenia due to lack of effort or to cogntive impairment? *German Journal of Psychiatry*.
Strauss, G. P., L. F. Morra, S. K. Sullivan, and J. M. Gold. 2014. The role of low cognitive effort and negative symptoms in neuropsychological impairment in schizophrenia. *Neuropsychology*.
Strutt, A. M., B. M. Scott, V. J. Lozano, P. G. Tieu, and S. Peery. 2012. Assessing sub-optimal performance with the test of memory malingering in spanish speaking patients with tbi. *Brain injury : [BI]* 26(6):853-863.
Stuss, D. T. 1995. A sensible approach to mild traumatic brain injury. *Neurology* 45(7):1251-1252.
Suhr, J. A. 2002. Malingering, coaching, and the serial position effect. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 17(1):69-77.

Suhr, J. A., and J. Gunstad. 2000. The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 15(5):415-424.
Sullivan, S., G. Strauss, K. Ossenfort, A. Katz, and J. Gold. 2014. C-83the role of iq in effort test performance in schizophrenia. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 29(6):604.
Swihart, A. A., K. M. Harris, and L. L. Hatcher. 2008. Inability of the rarely missed index to identify simulated malingering under more realistic assessment conditions. *Journal of clinical and experimental neuropsychology* 30(1):120-126.
Tan, J. E., D. J. Slick, E. Strauss, and D. F. Hultsch. 2002. How'd they do it? Malingering strategies on symptom validity tests. *The Clinical neuropsychologist* 16(4):495-505.
Tarescavage, A. M., D. B. Wygant, R. O. Gervais, and Y. S. Ben-Porath. 2013. Association between the mmpi-2 restructured form (mmpi-2-rf) and malingered neurocognitive dysfunction among non-head injury disability claimants. *The Clinical neuropsychologist* 27(2):313-335.
Temple, R. O., A. M. McBride, M. D. David Horner, and R. M. Taylor. 2003. Personality characteristics of patients showing suboptimal cognitive effort. *The Clinical neuropsychologist* 17(3):402-409.
Trueblood, W., and M. Schmidt. 1993. Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of clinical and experimental neuropsychology* 15(4):578-590.
Van der Does, A., and R. J. Van den Bosch. 1992. What determines wisconsin card sorting performance in schizophrenia? . *Clinical Psychology Review* 12:567 - 583.
Van Dyke, S. A., S. R. Millis, B. N. Axelrod, and R. A. Hanks. 2013. Assessing effort: Differentiating performance and symptom validity. *The Clinical neuropsychologist* 27(8):1234-1246.
Vanderploeg, R. D., and G. Curtiss. 2001. Malingering assessment: Evaluation of validity of performance. *NeuroRehabilitation* 16(4):245-251.
Vilar-Lopez, R., S. Santiago-Ramajo, M. Gomez-Rio, A. Verdejo-Garcia, J. M. Llamas, and M. Perez-Garcia. 2007. Detection of malingering in a spanish population using three specific malingering tests. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 22(3):379-388.
Walter, J., J. Morris, A. Swier-Vosnos, and N. Pliskin. 2014. Effects of severity of dementia on a symptom validity measure. *The Clinical neuropsychologist*:1-12.
Warrington, E. K. 1974. Deficient recognition memory in organic amnesia. *Cortex; a journal devoted to the study of the nervous system and behavior* 10(3):289-291.
Warrington, E. K., and A. M. Taylor. 1973. Immediate memory for faces: Long- or short-term memory? *The Quarterly journal of experimental psychology* 25(3):316-322.
Wasyliw, O. E., and J. L. Cavanaugh, Jr. 1989. Simulation of brain damage: Assessment and decision rules. *The Bulletin of the American Academy of Psychiatry and the Law* 17(4):373-386.
Wasyliw, O. E., L. S. Grossman, T. W. Haywood, and J. L. Cavanaugh, Jr. 1988. The detection of malingering in criminal forensic groups: Mmpi validity scales. *Journal of personality assessment* 52(2):321-333.
Webb, J. W., J. Batchelor, S. Meares, A. Taylor, and N. V. Marsh. 2012. Effort test failure: Toward a predictive model. *The Clinical neuropsychologist* 26(8):1377-1396.

Whitney, K. A., J. N. Hook, A. R. Steiner, P. H. Shepard, and S. Callaway. 2008. Is the rey 15-item memory test ii (rey ii) a valid symptom validity test?: Comparison with the tomm. *Applied neuropsychology* 15(4):287-292.

Williams, J. M. 2011. The malingering factor. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 26(3):280-285.

Williamson, D. J., M. Holsman, N. Chaytor, J. W. Miller, and D. L. Drane. 2012. Abuse, not financial incentive, predicts non-credible cognitive performance in patients with psychogenic non-epileptic seizures. *The Clinical neuropsychologist* 26(4):588-598.

Wu, T. C., M. D. Allen, N. J. Goodrich-Hunsaker, R. O. Hopkins, and E. D. Bigler. 2010. Functional neuroimaging of symptom validity testing in traumatic brain injury. *Psychological Injury and Law* 3:50 - 62.

Wygant, D. B., M. Sellbom, R. O. Gervais, Y. S. Ben-Porath, K. P. Stafford, D. B. Freeman, and R. L. Heilbronner. 2010. Further validation of the mmpi-2 and mmpi-2-rf response bias scale: Findings from disability and criminal forensic settings. *Psychological assessment* 22(4):745-756.

Yang, C. C., C. J. Kao, T. W. Cheng, W. H. Wang, R. L. Yu, Y. H. Hsu, and M. S. Hua. 2012. Cross-cultural effect on suboptimal effort detection: An example of the digit span subtest of the wais-iii in taiwan. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 27(8):869-878.

Young, G. 2014. Resource material for ehtical psychological assessment of symptom and performance validity, including malingering. *Psychological Injury and Law* 7:206 - 235.

Youngjohn, J. R., P. R. Lees-Haley, and L. M. Binder. 1999. Comment: Warning malingerers produces more sophisticated malingering. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists* 14(6):511-515.