Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2015-06-01

Passing or Failing of Symptom Validity Tests in Academic Accessibility Populations: Neuropsychological Assessment of "Near-Pass" Patients

Thomas Jeffrey Farrer
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Psychology Commons

BYU ScholarsArchive Citation

Farrer, Thomas Jeffrey, "Passing or Failing of Symptom Validity Tests in Academic Accessibility Populations: Neuropsychological Assessment of "Near-Pass" Patients" (2015). *All Theses and Dissertations*. 5944.
https://scholarsarchive.byu.edu/etd/5944

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Passing or Failing of Symptom Validity Tests in Academic Accessibility Populations:

Neuropsychological Assessment of "Near-Pass" Patients

Thomas Jeffrey Farrer

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Erin D. Bigler, Chair
Ramona O. Hopkins
Dawson W. Hedges
Joseph Olsen
Michael P. Brooks

Department of Psychology

Brigham Young University

June 2015

Copyright © 2015 Thomas J. Farrer

All Rights Reserved

ABSTRACT

Passing or Failing of Symptom Validity Tests in Academic Accessibility Populations: Neuropsychological Assessment of "Near-Pass" Patients

Thomas Jeffrey Farrer
Department of Psychology, BYU
Doctor of Philosophy

There is overwhelming evidence that the presence of secondary gain is an independent predictor of both performance validity and neuropsychological test outcomes. In addition, studies have demonstrated that genuine cognitive and/or psychological conditions can influence performance validity testing, both in the presence and absence of secondary gain. However, few studies have examined these factors in a large sample of academic accommodation seeking college students. The current study examined base rates of symptom validity test failure, the possibility of a "Near-Pass" intermediate group on symptom validity tests, the influence of diagnoses on performance indicators, and whether performance validity differed for "Near-Pass" patients relative to those who pass and those who fail performance validity indicators.

Keywords: symptom validity test, performance validity, academic accessibility, neuropsychological functioning

ACKNOWLEDGMENTS

I wish to thank my dissertation committee in training and supervision of this project. I also wish to acknowledge those who supported this project from BYU's Accessibility Center through data extraction and entry. I would like to acknowledge and thank the assistance of Cynthia Perry of BYU's Psychology Department and Dr. David Barker, Assistant Professor of Psychiatry and Human Behavior at Brown University and Rhode Island Hospital, who both provided valuable assistance on statistical analyses for this project. I wish to thank Drs. Stephen Correia and Donald Labbe of the Providence VA Medical Center for technical support for my dissertation defense. Finally, I would like to thank my wife, Loralee, and our two wonderful children, Evelyn and Vivian, for their constant love and support.

TABLE OF CONTENTS

Introduction 1

Assessment of Malingering Using SVTs 2

Operational Definition of Malingering 3

Slick Criteria of Malingering 4

The Fallacy of a True Cut-Score 7

SVTs and Cognitive Test Performance in Academic Populations 13

Method 15

Results 19

Discussion 42

Prevalence of SVT Failure 42

Latent Groups and a "Gray Zone" in SVT Outcomes 47

Cognition Functioning and SVT Outcomes 51

Implications of Failed SVTs on Clinical Data 55

Strengths and Limitations 60

Conclusions and Future Directions 62

References 67

Appendices 79

LIST OF FIGURES

Figure 1. Tri-modal Distribution Representing “Near-Pass” patients 13

Figure 2. Frequencies of Diagnostic Groups 24

LIST OF TABLES

Table 1. Overall Sample Descriptive Statistics 20

Table 2. Failure Rate of SVTs 21

Table 3. Model Fit Indices for 1- to 5- class solutions of SVTs 22

Table 4. Mean difference between those with diagnosis and those without 23

Table 5. Distribution of diagnosis in total sample 25

Table 6. Chi-Square analysis of diagnosis status across SVT outcome groups 26

Table 7. Chi-Square analysis of diagnosis status across Pass vs. Fail groups 27

Table 8. SVT difference across diagnostic classification 29

Table 9. Frequency table for Pass, Near Pass, and Fail groups for SVTs 31

Table 10. ANOVA for IQ testing in groups defined by TOMM trail 1.....33

Table 11. ANOVA for WJ-III in groups defined by TOMM trail 1.....35

Table 12. ANOVA for IQ testing in groups defined from RDS.....37

Table 13. Bonferroni correction for WAIS-IV across groups defined by RDS.....38

Table 14. ANOVA for WJ-III testing across groups defined by RDS 39

Table 15. : Bonferroni correction for WJ-III across groups defined by RDS.....40

Table 16. Above average performances with suboptimal SVT performance.....59

Passing or Failing of Symptom Validity Tests in Academic Accessibility Populations: Neuropsychological Assessment of "Near-Pass" Patients

The detection of malingering is a significant concern in the field of neuropsychology. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) defines malingering as an intentional production of false or exaggerated symptoms for secondary gain (American Psychological Association, 2013). Malingering has become a major area of assessment in clinical settings considering its putatively high occurrence. In the field of neuropsychology, base rates of malingering range from 8% in medical cases to 30% in disability cases (Mittenberg, Patton, Canyock, & Condit, 2002) and as high as 40% in clinical cases (Larrabee, 2003). Binder (1993) postulated that up to 33% of mild head injury cases malinger. In addition, Sollman and Berry (2011) suggests that these base rates are likely underestimated. In assessment of college students seeking academic accommodations, several studies likewise suggest that base rates of malingering are considerably high. One study found that 22% of college students feigned ADHD symptoms, and another study found that as many as 50% of college students fake ADHD symptoms (Sullivan, May, & Galbally, 2007; Marshall et al., 2010). A recent study suggests that in asymptomatic undergraduates not seeking secondary gain, suboptimal effort, as measured by three standard symptom validity tests, was as high as 55% (An, Zakzanis, & Joordens, 2013). It should be said, however, that recent work has challenged these numbers with estimates around 2% in non-clinical samples (Silk-Eglit, Stenclik, Gavett, Adams, Lynch, & Mccaffrey, 2014).

Considering multiple reports of high base rates, it is not surprising that research estimates the medico-legal financial cost of malingering to be between 5 and 20 billion dollars annually (Gouvier, Lees-Haley, & Hammer, 2003; Chafetz & Underhill, 2013). As a result, neuropsychologists are increasingly requested to assess patients in personal injury litigation,

worker's compensation settings, academic accommodation settings, and criminal culpability cases (Bianchini, Mathias, & Greve, 2001).

It is partially for the above reasons that multiple tests have been developed in an attempt to measure poor effort (Bianchini, Mathias, & Greve, 2001). Such tests are referred to as symptom validity tests (SVTs), or performance validity tests (PVT) (Bigler, 2014), and are designed to determine if individuals are putting forth adequate effort during neuropsychological testing. Note that although historically these tests have been referred to as SVTs and are often used interchangeably with PVTs, there is a push now to distinguish SVTs from PVTs. Performance validity tests are thought to reflect the effort put forth on objective measures of maximal cognitive ability, whereas SVTs are thought to reflect accuracy of symptom reporting on self-report measures of impairment (Larrabee, 2012; Van Dyke, Millis, Axelrod, & Hanks, 2013; Bigler, 2014; Institute of Medicine, 2015). However, for the purposes of this paper, the SVT categorization will be used throughout and used to refer to both SVT and PVT assessment.

One weakness of the proliferation of SVTs is that there are currently no guidelines for which SVTs are best or what the minimal objective level of "effort" is for diagnosing malingering (Bigler, 2011). There is also contradictory evidence to the high base rates noted above in that some literature suggests that SVTs have satisfactory psychometric properties in assessment of malingering in academic accommodation seeking college students (Jasinski, et al., 2011). As such, further research is needed to describe the sensitivity and specificity of various SVTs.

Assessment of Malingering Using SVTs

From the information of the prevalence of malingering discussed above, it is no surprise that symptom validity measures have proliferated in the field of clinical neuropsychology.

However, the type of measure used to assess malingering varies. Mittenberg et al. (2002) surveyed practicing clinicians and found that forced-choice measures were used in 57 percent of malingering cases and validity scales embedded within other tests were used in 38 percent of cases. The forced-choice method of assessing effort is based on tests of memory that appear to be difficult but are designed to actually be quite easy and can be passed with minimal cognitive effort. In fact, the assertion is that unless an individual has severe neurological damage, they should always score above chance (Larrabee, 2012) although this has been refuted in several studies (see discussion below). The embedded tests of effort are those that detect inconsistent symptom reporting. An example is the Fake Bad (F subscales) of the Minnesota Multiphasic Personality Inventory (MMPI; Graham, 2011). In neuropsychology, standalone SVTs are more often used because patients being evaluated are typically those complaining of cognitive changes more than personality and emotional changes and as neuropsychologists assess cognitive functioning in these patients, there arises a need to verify that these symptoms are not being exaggerated. As such, there is a need for an operational definition or criterion for what constitutes malingering over genuine effort.

Operational Definition of Malingering

As mentioned above, malingering can be defined as an intentional production of exaggerated symptoms for secondary gain (American Psychological Association, 2013). However, one limitation to the use of the term as defined by the APA is that the DSM-5 currently offers no diagnostic criteria for malingering. Specifically, the DSM limits malingering to a V-code in which malingering is not necessarily a mental disorder but that it might be important for the clinician to consider in the treatment of a patient. As a result, several studies have proposed systematic diagnostic criteria for describing malingering. One of the most often

cited studies is that of Slick, Sherman, and Iverson (1999) wherein the authors offer diagnostic criteria for *possible*, *probable*, and *definite* malingering of cognitive dysfunction.

Slick criteria of malingering. Slick et al. (1999) define malingering as "volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility." (p. 522). To further delineate the nature of malingering, these authors propose that there are degrees or magnitudes of malingering, and they describe three possible categories; possible, probable, and definite. In an attempt to make classification as objective as possible, the authors suggest using systematic criteria to determine malingering classification. The criteria, directly pulled from Slick et al. (1999, pp. 552-555), are as follows:

- Criterion A: Presence of a substantial external incentive – There must be at least one clearly identifiable and external incentive for exaggeration or fabrication at the time of testing.
- Criterion B: Evidence from neuropsychological testing – There must be evidence from testing as demonstrated from at least one of the following.

    1. Definite negative response bias – Below chance performance on one or more forced-choice measures of cognitive function.
    2. Probable response bias – Performance on one or more well-validated psychometric tests or indices designed to measure exaggeration or fabrication of cognitive deficits is consistent with feigning.
    3. Discrepancy between test data and known patterns of brain functioning – A pattern of neuropsychological test performance that is markedly discrepant from currently accepted models of normal and abnormal central nervous system (CNS) function. The discrepancy must be consistent with an attempt to exaggerate or fabricate neuropsychological dysfunction.
    4. Discrepancy between test data and observed behavior – Performance on two or more neuropsychological tests within a domain are discrepant with observed level of cognitive function in a way that suggests exaggeration or fabrication of dysfunction.
    5. Discrepancy between test data and reliable collateral reports – Performance on two or more neuropsychological tests within a domain are discrepant with day-to-day level of cognitive function described by at least one reliable collateral informant in a way that suggests exaggeration or fabrication of dysfunction.

6. Discrepancy between test data and documented background history – Improbably poor performance on two or more standardized tests of cognitive function within a specific domain (e.g., memory) that is inconsistent with documented neurological or psychiatric history.

- Criterion C: Evidence from self-report – The following are indicators of possible malingering and the presence of one or more of these provides additional support for a diagnosis of malingering.

    1. Self-reported history is discrepant with documented history – Reported history is different from documented medical or psychological history which suggests attempts to exaggerate or deny symptoms.
    2. Self-reported symptoms are discrepant with known patterns of brain functioning – Reported or endorsed symptoms are improbable in number, pattern, or severity or are inconsistent with expectations for the type or severity of documented injury or pathology.
    3. Self-reported symptoms are discrepant with behavioral observations – Reported symptoms are different from observed behavior.
    4. Self-reported symptoms are discrepant with information obtained from collateral informants – Reported symptoms, history, and behaviors are inconsistent with information provided by reliable informants.
    5. Evidence of exaggerated or fabricated psychological dysfunction – Self-reported symptoms are contradicted by behavioral observations or collateral information and well-validated validity scales are suggestive of exaggerated or feigned symptoms.

- Criterion D: Behaviors meeting necessary criteria from groups B or C are not fully accounted for by psychiatric, neurological, or developmental factors. The individual's behaviors are the product of informed, rational, and volitional effort aimed at acquiring external incentives.

Based on the above criteria, a diagnosis of definite malingering is put forth if the patient's behaviors occur in the presence of substantial external incentive (criterion A), if there is a negative response bias (criterion B1), and if the behaviors are not fully accounted for by psychiatric, neurological, or developmental reasons (criterion D). Probable malingering is diagnosed when the patient again meets criterion A, as well as two or more items from B2-B6 or C1-C5. Lastly, a diagnosis of possible malingering occurs when the patient again has external incentive (criterion A), meets one or more criteria from C1-C5 but not because of psychiatric, neurological, or developmental factors. Possible malingering is also considered when criteria for

definite or probably malingering are met except for criterion D, in which case, alternate etiologies should be specified.

More recent position papers provide additional recommendations to clinicians attempting to assess malingering. Bush et al. (2005) offer the following considerations which are more or less reiterated in a 2009 AACN consensus statement of malingering (Heilbronner et al., 2009): There must be consistency in self-reported data with presented symptoms; Clinicians need to consider neuropsychological and psychological test performance; and Clinicians also need to consider SVTs as well as forced-choice tests.

No consensus statement to date recommends using cut-scores as the absolute measure of effort. Although the above diagnostic criteria provide a useful way to assess malingering, Slick et al. also suggest that there are still significant limitations to malingering assessment. One such limitation is that the tests to assess malingering (SVTs) have less than perfect psychometric properties. More recent literature supports this assertion. For example, in a letter to the editor, Bigler (2011) suggests that "Often good versus poor effort is based on whether the selected SVT is performed above ("good effort") or below ("poor effort") an established cut-score. However, cut-scores create binary classifications with inherent limitations." (p. 751). In the AACN consensus statement cited above, we read the following: "Whether using a multivariable composite or a single test, neuropsychologists should not rely on single, fixed cut scores. Neuropsychologists appreciate and consider a range of cut scores and associated diagnostic test statistics in choosing the cut score to be applied to a specific case. The decision-making process occurs in different contexts, such that the relative costs of false positive and false negative errors will not be constant across situations." The authors continue, "To assist clinicians in this decision-making process, investigators, journal editors, and test publishers are strongly

encouraged to provide a broad range of cut scores with their respective diagnostic test statistics (e.g., sensitivities, specificities, and likelihood ratios).” (Heilbronner et al., 2009, p. 27).

The Fallacy of a True Cut-Score

As pointed out by Dwyer (1996), all cut-scores require judgment, and “Setting cut scores will invariably lead to errors in the desired classification scheme. Cut scores almost always impose external differentiations on a continuous distribution. Very few tests can distinguish reliably between examinees with adjacent scores, yet applying a cut score in effect forces such a distinction. No "true" cut score exists that will be found with the application of the correct method or a large enough sample of judges.” (p.360-361).

The acceptance of cut-scores has waxed and waned over the years but there has always been a concern that cut-scores are arbitrary and capricious. It is recommended by several authors to “avoid setting cut scores whenever possible” (Zieky, 2001, p. 24-25). The literature on this topic consistently points to human judgment being a major weakness of using cut-scores. Dwyer (1996) likens establishing cut-scores to establishing population parameters in statistics. For example, when trying to establish what the mean age is in a population of 10,000, using a sample of 9,000 should result in a sample mean closer to the true mean than if a sample of 90 was used. Likewise, a majority of methods for establishing cut-scores employ the use of judges, and thus using a larger sample of better trained judges would reduce some measurement error in establishing the cut-score. However, Dwyer maintains that because judgment is always involved, it is a mistake to assume that a cut score established by a larger group of people is in any way better than that previously established (Dwyer, 1996). Similarly, Zieky (2001) reports that, “The sense that there was a “correct” cutscore that could be established if all of the necessary resources were available, if an appropriate methodology were used, and if everything were done

correctly has generally been replaced with the knowledge that cutscores depend on values. As long as different people have different values, disagreements about cutscores will continue." (p. 45).

The idea that there is no true cut-score and that finding and establishing one is meaningless has been voiced by other authors as well. Cizek (1993) points out that most methods of establishing cut-scores use judges and that these judges (and their personal biases) will have to propose how a hypothetical individual will perform. Thus, establishing cut-scores is based off of "unobservable characteristics of hypothetical examinees and on untestable predictions about their performance." (p. 99). The author maintains that with such a method in place, finding a true passing score is absurd (Cizek, 1993). Similarly, Kane (1994) points out that with a different judge, time frame, and circumstance, cut-scores will change and that the cut-score can change dramatically without violating any principle related to the methods used to establish the cut score. "We create the standard; there is no gold standard for us to find, and the choices we make about where to set the standard are matters of judgment." (p. 427).

In a review of the methodological progress that has been made in the area of standard-setting, Zieky (2001) said, "The issue of finding the appropriate balance between passing those who should fail and failing those who should pass has continued to haunt people involved in setting cut scores. Regardless of improvements in methodology over the centuries, deciding what is appropriate remains very much a matter of subjective judgment." (p. 21).

Because of the subjective nature of setting cut-scores, caution should be used about what such cut-scores truly mean in clinical decision making and in social policy. When scores are at the extreme ends of a spectrum, little concern is given as to whether the person meets criteria for a passing score (Bigler, 2012). For example, the Test of Memory Malingering (TOMM) has a

maximum score of 50 with a cut-score of 45. Therefore, if individuals score well below chance, (e.g., 18), a clinician can have greater confidence in claiming the patient is malingering, or putting forth poor effort. This becomes more difficult when individuals are close to a cut-score. In other words, if an individual obtains a score of 44, he or she is said to be below the cut-score and is therefore thought to be exhibiting reduced effort. However, there are some studies that suggest that such close scores should not be used as an all-or-nothing measure of malingering. For example, Willis, Farrer, and Bigler (2011) demonstrated that even mild cognitive impairments can influence the results of effort tests that are thought to only be sensitive to extreme cognitive deficits. Bigler (2012) suggests that patients who are just below the cut-score can be thought of as "Near-Pass" SVT patients. He continues by demonstrating that "Near-Pass" patients can have clear and significant structural brain damage. One study demonstrated that individuals with diagnosable cerebral diseases and patients specifically with Alzheimer's disease frequently fall below SVT cut-scores (Merten, Bossink, & Schmand, 2007). In a similar study, 9.7-percent of Mild Cognitive Impairment (MCI) and 21.4-percent of moderate to severe dementia patients fell just below the cut score of TOMM trial 2 (Walter, Morris, Swier-Vosnos, & Pliskin, 2014). Similarly, patients with Huntington's disease are more likely to fail SVTs as their symptoms increase (Sieck, Smith, Duff, Paulsen & Beglinger, 2013). Neuropsychiatric patients are also likely to fall below cut-scores of SVTs (Gorissen, Sanz, & Schmand, 2005). In fact, there are multiple studies demonstrating that patients with well-documented conditions commonly fail SVTs (Greve, Ord, Curtis, Bianchini, & Brennan, 2008; Howe & Loring, 2009; Keary et al., 2013; Eichstaedt et al., 2014). Of importance to the present investigation is the fact that no study to date has determined the existence of a "Near-Pass" group in accommodation seeking college students with legitimate deficits. However, to support the above literature that

genuine cerebral damage modifies SVT performance, a double blind cross over study found that normal patients taking lorazapam, a common benzodiazepine used to treat anxiety, had decreased performance on the Word Memory Test and then pass this SVT when not on medication (Loring et al., 2011). This accumulated evidence suggests that false-positives, or the misclassification of truly impaired patients as malingering, are possibly more likely than previously assumed. Indeed, Bigler (2012) suggests that ignoring the evidence that neurologically impaired individuals may legitimately fail SVTs increases the risk of a Type II error. This is further supported by recent work illustrating that the cognitive load (i.e., systematically adding distractions during testing) increase the likelihood that individuals with neurological diagnoses fail SVTs (Leighton, Weinborn, & Maybery, 2014). In addition, as noted in the above consensus statements, the use of more than one SVT is recommended in determination of effort during testing. However, false-positive rates become more problematic as the number of SVTs used increases. A recent study utilizing a Monte-Carlo simulation demonstrated that the false-positive rate increases in relation to the number of effort indicators used (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013).

The error rates that surround cut-scores constitute an understudied area in the literature. In addition, in an attempt to identify feigned ADHD in an accommodation seeking college population, Young and Gross (2011) found that embedded effort measures of the MMPI-2 have poor sensitivity and specificity and erroneous cut-scores for detecting malingering. These findings support the literature that there is no true cut-score and that relying on a cut-score as an all-or-nothing measure of poor effort is clinically irresponsible.

As mentioned above, Binder (1993) postulated that up to 33% of mild head injury patients in litigation malinger while additional studies found that 22-50% of college students

fake ADHD symptoms while being assessed for academic accommodations (Marshall et al., 2010; Sullivan et al., 2007). In other words, the potential for secondary gain apparently increases the chance of feigned impairment. However, few studies have examined the base rate of falling below cut-scores in legitimate populations that ostensibly have no secondary gain to malinger. Those that have suggest that a significant portion of patients can score below cut-scores, even without the secondary motivation to malinger. For example, Kirkwood and Kirk (2007) consecutively studied 193 cases of children and adolescents with a history of mTBI and found that 33 of these individuals failed the Medical Symptom Validity Test. Of these, only one was found to be related to litigation status while the remainder failed the MSVT without the presence of secondary gain or motivation to malinger. Loring, Lee, and Meador (2005) likewise found that 20 percent of epilepsy surgery candidates without external incentive are classified as either questionably valid or invalid on the Victoria Symptom Validity Test (VSVT). A case study by Goodrich-Hunsaker and Hopkins (2009) revealed that non-litigating amnestic patients with significant hippocampal damage may in fact pass most subtests of the Word Memory Test but also demonstrate profoundly impaired free recall. Again, as mentioned above, no study to date has examined these effects in college students with legitimate cognitive and psychological deficits such as learning disorders or ADHD, although some research refutes that ADHD symptoms produce reduced SVT performance (Harrison, Flaro, & Armstrong, 2014).

Additional literature also suggests that stand-alone SVTs may not be ideal for identifying malingering and may, in fact, falsely identify patients as malingering. Bush et al. (2005) suggest that formal tests are not the only way to measure malingering and that the assessment of malingering also includes clinical observation. In addition, Bush et al. point out that SVTs that fall slightly below cut-off may not imply malingering and that additional indicators are needed

first. However, consensus groups in the field of neuropsychology recommend the use of SVTs in forensic settings (Bush et al., 2005; Heilbronner et al., 2009). The main problem with this is that the use of cut-scores ignores the psychometric phenomenon of measurement error. Proponents of SVTs maintain that measurement error and confidence intervals are not critical for SVTs because the tests are intentionally designed to be easy and can therefore be dichotomized as pass or fail (Green & Flaro, 2003; Sweet, 1999; Carone, 2014). As reviewed by Leighton et al., (2014), recognition memory is typically thought of as an "automatic process" (p. 880) but when the cognitive load is increased during the learning phase of SVTs, performance declines. For example, one study examined the pass-fail rate of the TOMM and the WMT among severe TBI patients and found that with standard administration, 44% and 16% scored below the WMT and TOMM cut-scores, respectively. However, when the learning phase includes a condition of increased cognitive load (i.e., adding distraction), the fail rates increased to 75% on the WMT and 33% on the TOMM (Batt, Shores, & Chekaluk, 2008). This same pattern of reduced SVT performance during increased cognitive loads is also seen in healthy controls with no incentive to malinger (Leighton et al., 2014). The suggestion that SVTs require cognitive capacity is supported by imaging studies. For example, functional brain imaging studies suggest that some SVTs actually act as measures of cognitive function and will therefore have variable outcome scores in assessment (Allen, Bigler, Larsen, Goodrich-Hunsaker, & Hopkins, 2007; Larsen, Allen, Bigler, Goodrich-Hunsaker, & Hopkins, 2010). In addition, a recent meta-analysis found that the overall sensitivity for common SVTs in neuropsychological assessment was .69 while the specificity was .90 (Sollman & Berry, 2011). Though these figures are as strong as other psychometric assessment tools, they imply that there will be some measurement error.

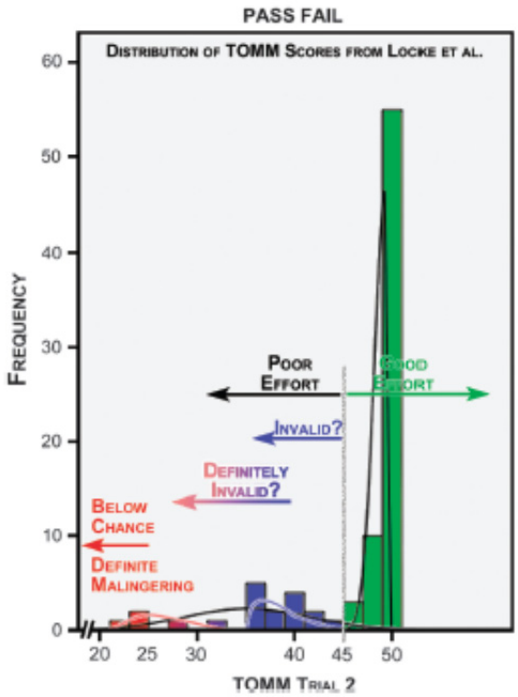

Figure 1. Tri-modal distribution of TOMM scores as noted in Bigler (2012) with those just below the cut-score of 45 being described as "Near-Pass" SVT patients, while those above 45 are good effort patients, and those below 30 are below chance patients. Bigler argues that 'Near-Pass" patients are qualitatively different that individuals in the other groups.

In summary, there fails to be a consensus about the use of SVT cut-scores in clinical settings. There is substantial literature that suggests scores on SVTs are more difficult to interpret when individuals score at or near the cut-score. Also, this seems to be a bigger problem when an individual does in fact have a legitimate condition that may have subtle effects on test performance, even without the chance of secondary gain.

SVTs and Cognitive Test Performance in Academic Populations

Secondary gain is a critical issue of SVT outcome research and secondary gain could certainly play a role for SVT outcomes for academic populations. For instance, those who acquire some form of accommodations may be given extended time on tests, prolonged due dates for assignments, private test environments, alternative courses, shortened workloads, and even

prescription medications (i.e., stimulants) for ADHD. Therefore, measuring effort in accommodation seeking populations is important.

As mentioned above, few studies have examined malingering in accommodation seeking college students and the neuropsychological functioning of this population in relation to SVT performance. Sullivan et al. (2007) retrospectively examined SVT and neuropsychological performance of 66 adults who were evaluated for ADHD, learning disorder (LD), or both. The authors used the Word Memory Test (WMT) as the SVT measure. They found that 47% of those being evaluated for ADHD failed the WMT, while 15% of the LD and 9% of the combined ADHD/LD participants failed the WMT. In addition, and critical for the present investigation, Sullivan et al. also found that those who failed the WMT had significantly lower scores on the WAIS-III and the CVLT. A more recent study attempted to replicate these findings using 85 adults being assessed for ADHD (Suhr, Hammers, Dobbis-Buckland, Zimak, & Hughes, 2008). Their findings support those of Sullivan et al. (2007) in that ADHD participants who failed the WMT had significantly lower memory and executive function performance compared to those who scored above the WMT cut-score. These findings were challenged by a recent review of malingered ADHD symptoms in college students, which suggested that neuropsychological profiles between malingered ADHD and valid ADHD are indistinguishable (Musso & Gouvier, 2014). In sum, these studies suggest that screening for effort in accommodation seeking students is important. However, no study has examined the above mentioned tri-modal distribution of scores in SVT performance or whether neuropsychological test performance differs between those who clearly pass, those who clearly fail, and the “Near-Pass” individuals described by Bigler (2012). Considering the evidence that those with legitimate and genuine neuropsychiatric conditions can fall just below cut-scores, it is critical to elucidate whether differences in

neuropsychological performance exist between these three groups. The present study will examine these differences in archived data sets of adult accommodation seeking college students.

Method

This dissertation sought to answer multiple questions. First, the present investigation assessed whether Bigler's (2012) "Near-Pass" group exists. In his 2012 paper, Bigler asserted that there are three latent distributions in a set of malingering scores; pass, fail, and "Near-Pass" (see Figure 1 above). However, in a rebuttal to Bigler's article, Larrabee (2012) denied such assertion suggesting there are only two groups; pass or fail. Here, a Latent Profile Analysis was used to determine if a two class distribution was better than one and if three was better than two, etc. Evaluation of the number of classes was based on multiple fit statistics. In other words, information criteria were used to measure the relative fit of different models with different number of classes (e.g., 3 compared to 2) (Lanza & Rhoades, 2013). The lowest values among these indices signify better overall model fit. The fit criteria included the Bayesian Information Criterion (BIC) (Schwarz, 1978), the Sample Size Adjusted Bayesian Information Criterion (SSABIC) (Sclove, 1987), and the Akaike Information Criterion (AIC) (Akaike, 1974). In addition, the Lo-Mendell-Rubin (LMR) adjusted likelihood ratio test was used to determine if the fit of any given model with k classes was a better fit than a k-1 or k+1 class model (Lo, Mendell, & Rubin, 2001). The use of the LMR test allows for statistical comparison between classes with a significant p-value ($p < .05$) indicating that a specified k class model represents a better fit than k-1 or k+1 (Nylund, Asparouhov, & Muthen, 2007). Model fit measurement was done iteratively from one class to five classes. All LPA analyses were conducted with Mplus 6.01 (Muthen BO, Muthen LC. *Mplus Version 6.01*. Los Angeles: Muthen and Muthen, 2011). For the present study, stand-alone tests providing performance validity scores from the WMT and the TOMM,

and an embedded measure, reliable digit span (RDS), from the WAIS-III & IV were used to measure whether Bigler's "Near-Pass" group was plausible (Bigler 2012).

Second, considering the literature suggesting that legitimate neurological and psychological deficits can result in below cut-score performance on SVTs, performance on such effort tests was examined across diagnoses with the hypothesis that individuals with a legitimate diagnosis were in the "Near-Pass" group more often than individuals with no diagnosis or individuals judged as outright malingerers. Third, no studies have examined cognitive functioning in relation to "Near-Pass" individuals compared to outright pass or outright fail individuals. The present study hypothesized that the "Near-Pass" individuals have significantly poorer cognitive functioning compared to the other two groups.

This study was reviewed and approved by the BYU Institutional Review Board for Human Subjects (See Appendix A). Data for this study came from an archived data base of approximately 540 undergraduate and graduate students who presented at the BYU Accessibility Center for psycho-educational evaluation to determine if they had any learning and/or neuropsychological disorders that would qualify them for academic accommodations. Participants represented both male and female BYU students between the ages of 18 and 60. At the time of testing, all individuals signed a release allowing their individual but de-identified data to be used for research (see appendix B). All data used in this study was anonymous. Data on malingering from the UAC was gathered on students assessed from 2007 to 2014. Each student underwent clinical interview and completed a series of neurocognitive tests to determine the presence or absence of a deficit that would require academic accommodations. Diagnoses were made based on both DSM-IV diagnostic classifications (American Psychological Association, 2000) and group consensus from BYU Accessibility Center clinicians. Clinicians included two

or more doctoral level clinical psychologists, a clinical neuropsychologist, and graduate level clinicians who were completing practica under supervision.

The archived data was used to determine the relative difference between three groups – those that outright passed SVTs, those that outright failed, and those that are described in the literature to be "Near-Pass". Considering the literature suggesting that individuals with legitimate psychological deficits frequently fail SVTs, standard cut-scores were used to determine group status. Under these groupings, an analysis of variance (ANOVA) was used to establish whether group differences are statistically meaningful across cognitive testing results. For the student academic accommodation seeking group, the following measures were used in assessment: the Wechsler Adult Intelligence Scale – 3rd edition (WAIS-III), the Wechsler Adult Intelligence Scale – 4th edition (WAIS-IV), the Test of Memory Malingering (TOMM), Green's Word Memory Test (WMT), The Delis-Kaplan Executive Function System (DKEFS), the California Verbal Learning Test – 2nd edition (CVLT-II), the Story Recall (immediate and delayed) subtest of the Woodcock-Johnson III Test of Achievement (WJ-III), the Wechsler Memory Scale – 3rd edition (WMS-III), and the Wechsler Memory Scale – 4th edition (WMS-IV).

Reliable Digit Span (RDS) is used as an embedded measure of test effort. This differs from overt measures of malingering such as the TOMM. This test uses a cut-score much like other symptom validity measures. A score of ≤ 7 is the cut-score for poor effort. Digits forward and backward, measures of a working memory subtest of the WAIS-IV, are first administered in the standardized fashion. The RDS is calculated by taking the maximum number of digits forward and adding this number to the maximum number of digits backward where the max has to come from a set where both trials were passed. A recent study used meta-analysis to calculate global sensitivity and specificity for RDS across multiple studies. This study found a global

sensitivity between 48-58% and a global specificity of 82-85% (Schreoder, Twumasi-Ankran, Baade, & Marshall, 2012). This study also suggests that a cut-score of ≤ 6 may be more appropriate depending on the group being tested. However, the traditional cut-score of ≤ 7 will be used in this study.

The TOMM is an SVT used to determine if individuals are putting forth adequate effort during testing. It is comprised of 50 simple line drawings that are presented to the examinee one at a time for approximately three seconds each. Following presentation, the examinee is then shown two pictures, one of which was one of the original 50 items along with one foil, and the examinee has to point to the one they saw before. This is repeated over two trials with the same 50 pictures presented but in an alternate order. Overall, the examinee receives a score out of 50 on each of the two trials. The test is designed with a score of 45 being the cut-score for adequate effort. Malingering is suspect with TOMM performances that are below chances on any trial, or when performance is below 45 on trial 2. However, numerous research studies with clinical and non-clinical samples have demonstrated that using trial 1 of the TOMM can have robust clinical utility (Gavett, O'Bryant, Fisher, & McCaffrey, 2005; O'Bryant, Engel, Kleiner, Vasterling, & Black, 2007; O'Bryant et al., 2008) . As such, for this study, scores below 45 on either trial 1 or 2 will be considered failing.

Green's Word Memory Test (Green, 2003) evaluates immediate and delayed recognition of 20 semantically related word pairs (Pig – Bacon). The list of words is presented twice (PC or read by examiner). The Immediate Recognition (IR) trial follows the presentation of the words. The examinee must then choose out of 40 new pairs which word was one of the original (Cow – Pig). Unknown to the examinee, there is then a 30 minute delay where the patient has to identify the 40 original words with new foils in the mix (Pig – Feed). The primary measure of

malingering is the IR, DR, and consistency between responses (CNS). Invalid performance on the WMT is determined with a cut-score where a clear fail occurs when any one of the IR, DR, CNS trials falls ≤ 82.5% correct.

The CVLT-II is a rote verbal memory test in which the examinee is read a list of words and asked to memorize them over repeated trials. It is often used as a measure of learning and immediate and delayed verbal memory (Delis, Kramer, Kaplan & Ober, 2000) but also has a forced-choice section that is often used as another measure of testing effort. The WMS-III and IV measure an individual's memory abilities based off seven subtests. It is composed of auditory, visual, working, immediate, and delayed memory indices (Wechsler, 2002). Memory abilities were also assessed via story memory recall of the Woodcock-Johnson III, Test of Achievement. This test includes immediate recall of multiple short stories followed by a delayed recall portion. Scores for this test were scaled on a standard score where the mean is 100 and the standard deviation is 15. The WAIS-III and IV are measures of intellectual functioning comprised of Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed indices. These four index scores were used in calculation of cognitive function in the present study. Note that given the time frame of data collection, some individuals were administered different versions of the same test (i.e., WAIS-III or WMS-III, while more recent assessments included WAIS-IV and WMS-IV).

Results

Overall sample characteristics are detailed in Table 1. A total of 540 individuals met inclusion criteria from the archival data set. The mean age was 24.9 (SD = 7.1) and the mean years of education at the time of assessment was 14.7 (SD = 1.4). Most of the sample was male (62%). Of the total sample, not all individuals were administered SVTs or the same SVT. Of

those that were administered an SVT, only a small subset scored below traditional cut-scores (see Table 2). Also, of those who were administered an SVT of any type, not all were administered every neuropsychological test. For instance, since this is archival data, some earlier participants were administered the WAIS-III or WMS-III while more recent assessments included WAIS-IV and WMS-IV. Accordingly, usable sample sizes for any given comparison are considerably smaller than expected (See analyses below for additional data on sample sizes per analysis). Table 2 details the rate of failure across the various SVTs administered and frequency of test administration (n of administrations ranged from 228 to 324).

The number of administrations differs between trial 1 and trail 2 of the TOMM possibly because it is often the case that if an individual obtains a perfect score on trial 1, clinicians frequently forego the second trial. The WMT appears to have been used about as often as TOMM. Note that RDS yielded the highest overall sample size. This was ostensibly due to the fact that a large portion of the accommodation seeking students were given IQ testing as part of these assessments, which includes the digit span subtest from which RDS is calculated. Also, analysis of failure rates for each SVT demonstrates that the mean score for each would be considered passing, suggesting that on average, this particular population is likely to pass SVT evaluation more often than previous literature suggested. In fact, only 2-percent (n = 4) failed (i.e., fell below the cut-score) TOMM trial 2 of 231 administrations.

Table 1: *Overall Sample Descriptive Statistics*

	N	Min	Max	Mean	Std. D
Age	540	18	60	24.93	7.11
YOE	538	12	20	14.67	1.40

Note. YOE = Years of Education; YOE was not reported in two cases.

Table 2: *Failure Rate of SVTs*

Test	N	Min	Max	Mean	Std. D	Below Cut-Score	% Fail
TOMM Trial 1	228	26	50	48.00	3.50	28	12 %
TOMM Trial 2	231	29	50	49.67	1.77	4	2 %
WMT IR	254	15	100	95.19	9.39	22	9 %
WMT DR	253	45	100	95.55	8.52	19	8 %
WMT CNS	252	15	100	93.31	11.04	28	11 %
RDS	324	4	16	9.07	2.01	67	21 %

Note. SVT = Symptom Validity Test; TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span.

Latent Profile Analysis models were compared using model fit indices. Model fit statistics are presented in Table 3. The level of entropy was acceptable across all classes (≥ .80). A 5-class solution achieved the lowest AIC, BIC, and SSABIC values suggesting five latent classes within the SVT data. However, results of the Lo-Mendell-Rubin test suggest that no individual class *k* yields a better fit than any other class, *k-1*or *k+1*. In other words, no class solution (*k*) was statistically better than another. Thus, no single class solution was identified as a reasonable fit for this data. Although the overall sample sizes for each SVT appear adequately large, it is important to note that not all individuals were administered the same SVTs. In other words, although 231 individuals were given trial 2 of the TOMM, the sample size of individuals that were given this test *and* all other SVTs became as low as to undermine the LPA results. Note too that the number of individuals who fell below traditional cut-scores was also fairly low (e.g., TOMM trial 2 had four individuals with scores below the cut-score). This suggests the possibility of a truncated range of scores on SVT measures, at least in this sample. Taken together, the results of the LPA are limited by low cross-over n's between each SVT and a truncated range for any given SVT.

Table 3: *Model Fit Indices for 1- to 5- Class Solutions of SVTs*

Classes	No. of Parameters	LL	AIC	BIC	SSABIC	LMR	*p*	Entropy
1	12	-4610.78	9245.56	9295.47	9257.38	-	-	-
2	19	-4248.49	8534.98	8613.99	8553.70	708.16	0.35	0.96
3	26	-4086.27	8224.55	8332.68	8250.16	317.07	0.71	0.87
4	33	-4013.46	8092.92	8230.17	8125.43	135.52	0.78	0.85
5	40	-3856.77	7793.53	7959.89	7832.94	183.25	0.26	0.80

Note. LL = Log-Likelihood; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; SSABIC = sample size adjusted BIC; LMR = Lo-Mendel-Rubin value; p-value is for the LMR test.

There were 429 (79%) individuals out of the whole sample that met criteria for a least one diagnosis. The total sample was dichotomized as either having no diagnosis or having at least one. The group with no diagnosis constitutes all individuals receiving no diagnosis following assessment. A t-test was used to determine if SVT performance varied according to diagnosis status (no diagnosis versus any diagnosis). Results are presented in Table 4 which indicate that TOMM trial 1 reached significance ($p = 0.05$) with individuals with any diagnosis scoring 1.02 points below those with no diagnosis. Also, one standard deviation below the mean for the group with any diagnosis places them below the cut-score whereas one standard deviation below the mean for the group without a diagnosis remains within the accepted pass parameters of TOMM trial 1. As indicated above, out of 231 administrations of TOMM trial 2, only four individuals fell below the cut-score. As such, mean differences between groups were not calculated and TOMM trail 2 data are not presented in Table 4. No statistical differences were observed for WMT IR or WMT DR. Marginal significance was observed for WMT CNS but in an unexpected direction with *better* mean performance for individuals having received a diagnosis. Lastly, RDS reached the highest level of significance ($p = .002$) with lower mean performance for those with a diagnosis compared to no diagnosis.

Table 4: *Mean Difference Between Those With Diagnosis and Those Without*

Test	n	M(SD)	t	df	*p*
TOMM Trial 1					
Any Diagnosis	166	47.72 (3.86)	1.96	226	0.050
No Diagnosis	62	48.74 (2.11)			
WMT IR					
Any Diagnosis	209	95.69 (9.03)	-1.68	252	0.064
No Diagnosis	45	92.83 (10.66)			
WMT DR					
Any Diagnosis	207	95.93 (7.60)	-1.72	250	0.086
No Diagnosis	45	93.50 (11.80)			
WMT CNS					
Any Diagnosis	207	93.93 (10.25)	-1.93	250	0.054
No Diagnosis	45	90.44 (13.88)			
RDS					
Any Diagnosis	279	8.9 (1.98)	3.10	321	0.002
No Diagnosis	44	9.93 (1.96)			

Note. TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span.

Next, Figure 2 and Table 5 below display the frequency of broad diagnostic classifications among the total sample. Because of the nuanced differences between sub-classifications, multiple conditions were combined. This allowed for simplified statistical exploration and increased sample size per classification. For instance, all individuals with one diagnosis of any type of learning disability (LD) were combined into an LD group. The same was true of those with any subtype of an ADHD diagnosis. There were, of course, many individuals with more than one diagnosis. All individuals with an LD *and* ADHD diagnosis were combined into a single ADHD+LD group. All individuals with a type of LD, ADHD, and an axis I or II affective or personality disorder classification were combined into a single ADHD/LD + Affective/PD group. The last category was more difficult to define – there were many individuals with neurological conditions (e.g., seizure disorders, TBI, & brain tumor) and

individuals diagnosed simply as “Cognitive Disorder NOS”. It was assumed that those with Cognitive Disorder NOS must not have met criteria for ADHD, LD, or any other type of axis I or II condition. As such, all individuals with a neurological disorder and all individuals with a diagnosis of Cognitive Disorder NOS were combined into a single group. There were 12 individuals that did not seem to fit any of the above classifications (e.g., Asperger’s syndrome, cerebral palsy, Developmental Coordination Disorder, various visual impairments, etc.). These individuals were excluded from the descriptive statistics below. Findings indicate that of those that were ultimately given a diagnosis following assessment, most met criteria for some type of LD or ADHD. In fact, LD and/or ADHD diagnoses accounted for almost 60-percent of all diagnoses. In the current investigation, it is not known why some individuals were not given a diagnosis after assessment. In other words, it is unclear if they did not meet criteria or if a diagnosis was withheld because they were sub-threshold on DSM-IV criteria.

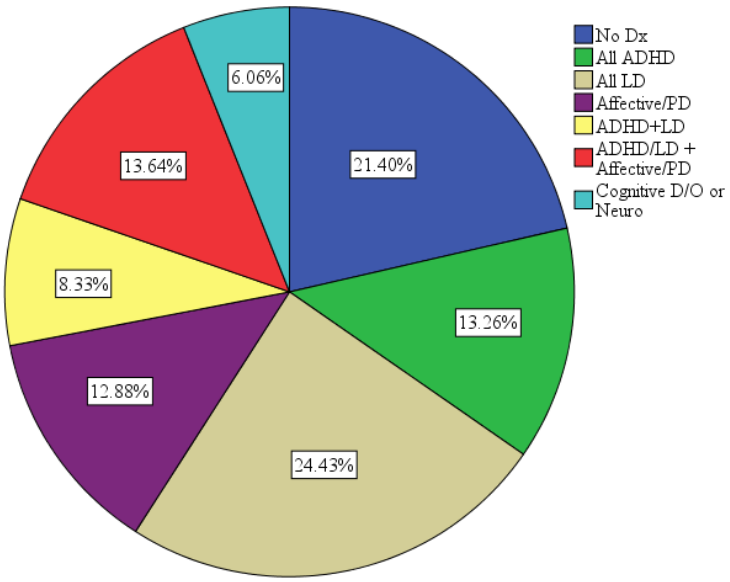

Figure 2. Distribution of diagnoses across entire sample. Dx = Diagnosis; ADHD = Attention Deficit/Hyperactivity Disorder; LD = Learning Disorder; PD = Personality Disorder; Cognitive D/O or Neuro = Cognitive Disorder or neurological disorder.

Table 5: *Distribution of Diagnosis in Total Sample*

Diagnostic Groups	Frequency	Percent
No Dx	113	21.4
All ADHD	70	13.3
All LD	129	24.4
Affective/PD	68	12.9
ADHD+LD	44	8.3
ADHD/LD + Affective/PD	72	13.6
Cognitive D/O or Neuro	32	6.1
Total	528	100

Note. Dx = Diagnosis; ADHD = Attention Deficit/Hyperactivity Disorder; LD = Learning Disorder; PD = Personality Disorder; Cognitive D/O or Neuro = Cognitive Disorder or neurological disorder.

Another aim of this study was to determine whether individuals with a diagnosis were more likely to fail (fall below cut-score) SVTs. This was assessed in two sets of chi-square analyses – a 2x3 (diagnosis and no-diagnosis by Pass, Fail, and Near Pass) and a 2x2 (diagnosis and no-diagnosis by Pass and Fail, where Fail is defined as scores below the established cut-score). Table 6 provides the results of the 2x3 chi-square analysis. Results suggest that expected counts and observed counts were statistically different for diagnostic classification by Pass, Fail, and Near Pass group status on WMT IR, WMT DR, and WMT CNS. However, there is a critical assumption of chi-square analysis which specifies that if more than 20% of the cells in a frequency table are ≤ 5, the chi-square cannot be interpreted reliably. This assumption was violated in all chi-square analyses except RDS, which had a non-significant *p*-value.

Table 6: *Chi-Square Analysis of Diagnosis Status Across SVT Outcome Groups*

		TOMM Trial 1						
		Pass	Near Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	57	4	1	62	2	3.03*	0.22
	Expected Count	54.4	3.8	3.8	62			
Any Dx	Count	143	10	13	166			
	Expected Count	145.6	10.2	10.2	166			
Total	Count	200	14	14	228			
	Expected Count	200	14	14	228			
		WMT IR						
		Pass	Near Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	36	4	5	45	2	9.12*	0.10
	Expected Count	41.1	2	2	45			
Any Dx	Count	195	7	6	208			
	Expected Count	189.9	9	9	208			
Total	Count	231	11	11	253			
	Expected Count	231	11	11	253			
		WMT DR						
		Pass	Near Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	38	1	6	45	2	8.9*	0.01
	Expected Count	41.6	1.3	2.1	45			
Any Dx	Count	195	6	6	207			
	Expected Count	191.4	5.8	9.9	207			
Total	Count	233	7	12	252			
	Expected Count	233	7	12	252			
		WMT CNS						
		Pass	Near Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	36	4	5	45	2	5.9*	0.052
	Expected Count	40.2	2.7	2.2	45			
Any Dx	Count	188	11	7	206			
	Expected Count	183.8	12.3	9.8	206			
Total	Count	224	12	12	251			
	Expected Count	224	12.0	12	251			
		RDS						
		Pass	Near Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	39	5	0	44	2	3.25	0.20

	Expected Count	34.8	7.9	1.2	44
Any Dx	Count	216	53	9	278
	Expected Count	220.2	50.1	7.8	278
Total	Count	255	58	9	322
	Expected Count	255	58	9	322

Notes: * Cells have an expected count less than 5, which violates assumptions of Chi-Square thus making results uninterruptable. Dx = Diagnosis; TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span.

The results of the 2x2 analysis are represented in Table 7 and demonstrate that expected counts and observed counts were statistically different for diagnostic classification by Pass and Fail group status on WMT IR, WMT DR, WMT CNS, and RDS. Again, there were low counts in WMT IR and WMT DR. The WMT CNS and RDS remained usable but yielded opposite findings with a higher observed count than expected of RDS failure among those with a diagnosis but a lower observed count than expected for WMT CNS failure among those with a diagnosis.

Table 7: *Chi-Square Analysis of Diagnosis Status Across Pass vs. Fail Groups*

		TOMM Trial 1					
		Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	57	5	62			
	Expected Count	54.4	7.6	62			
Any Dx	Count	143	23	166	2	1.4	0.24
	Expected Count	145.6	20.4	166			
Total	Count	200	28	228			
	Expected Count	200	28	228			
		WMT IR					
		Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	36	9	45			
	Expected Count	41.1	3.9	45			
Any Dx	Count	196	13	209		8.9*	0.00
	Expected Count	190.9	18.1	209			
Total	Count	232	22	254			
	Expected Count	232	22	254			

		WMT DR					
		Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	38	7	45			
	Expected Count	41.6	3.4	45			
Any Dx	Count	195	12	207		5.0*	0.02
	Expected Count	191.4	15.6	207			
Total	Count	233	19	252			
	Expected Count	233	19	252			
		WMT CNS					
		Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	36	9	45			
	Expected Count	40	5	45			
Any Dx	Count	188	19	207	2	4.4	0.03
	Expected Count	184	23	207			
Total	Count	224	28	252			
	Expected Count	224	28	252			
		RDS					
		Pass	Fail	Total	df	χ^2	*p*
No Dx	Count	39	5	44			
	Expected Count	34.9	9.1	44			
Any Dx	Count	217	62	279	2	2.7	0.09
	Expected Count	221.1	57.9	279			
Total	Count	256	67	323			
	Expected Count	256	67	323			

Notes: * Cells have an expected count less than 5, which violates assumptions of Chi-Square thus making results uninterruptable. Dx = Diagnosis; TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span.

Table 8 represents an analysis of variance of SVTs across these diagnostic classifications. No comparison was statistically meaningful across groups. A post-hoc analysis was also run in order to determine if individual groups differed from each other on any given SVT. No statistical differences were noted (table for this analysis is presented in Appendix C). Regarding Table 8, although no statistical differences were noted in mean SVT performance across diagnostic classifications, a few interesting features emerged. First, for TOMM trial 1, all mean performances were in the passing range (above a cut-score of 45). Only the ADHD group

recorded a performance close to chance (26/50). On TOMM trial 2, there was a trend-level significance (p = .06) with the lowest mean performance recorded for the Affective/PD group and near perfect performances for all other groups. Again, only the ADHD group recorded a performance near chance (29/50). However, the mean performances for TOMM trial 2 were all in the passing range. The results of the RDS were somewhat surprising given the fact that RDS is based on the digit span subtest of the WAIS III or IV. Here, it was assumed that since digit span is a test of working memory and attention, those with putative attentional impairments would have lower performance. Again, there was no statistical difference between groups, but those with ADHD performed at a mean of 9.5, which is the highest recorded mean across diagnostic groups.

Table 8: *SVT Difference Across Diagnostic Classification*

						95% CI					
		N	Mean	SD	SE	Lower	Upper	Min	Max	F	*p*
TOMM Trial 1	All ADHD	24	47.3	5.2	1.1	45.2	49.5	26	50	1.266	.28
	All LD	44	48.6	2.2	0.3	48.0	49.3	40	50		
	Affective/PD	28	46.7	5.7	1.1	44.5	48.9	31	50		
	ADHD+LD	16	48.1	3.0	0.7	46.5	49.6	40	50		
	ADHD/LD + Affective/PD	31	48.2	2.3	0.4	47.3	49.0	41	50		
	Cognitive D/O or Neuro	23	46.9	4.1	0.9	45.1	48.7	36	50		
	Total	166	47.7	3.9	0.3	47.1	48.3	26	50		
TOMM Trial 2	All ADHD	25	49.0	4.3	0.9	47.2	50.7	29	50	2.145	.06
	All LD	43	50.0	1.4	0.2	49.8	50.6	49	50		
	Affective/PD	28	48.8	2.7	0.5	47.7	49.8	41	50		
	ADHD+LD	18	50.0	0.0	0.0	50.0	50.0	50	50		
	ADHD/LD + Affective/PD	32	49.9	0.4	0.1	49.8	50.0	48	50		
	Cognitive D/O or Neuro	23	49.5	1.2	0.3	48.9	50.0	45	50		
	Total	169	49.6	2.2	0.2	49.3	49.9	29	50		
WMT IR	All ADHD	26	93.7	10.5	2.1	89.4	97.9	53	100	.975	.43
	All LD	66	97.5	4.6	0.6	96.3	98.6	75	100		
	Affective/PD	31	94.4	11.1	2.0	90.4	98.5	50	100		

	ADHD+LD	20	96.9	3.2	0.7	95.4	98.4	88	100		
	ADHD/LD + Affective/PD	51	95.3	12.6	1.8	91.7	98.8	15	100		
	Cognitive D/O or Neuro	11	95.2	5.6	1.7	91.4	99.0	83	100		
	Total	205	95.8	9.1	0.6	94.6	97.1	15	100		
WMT DR	All ADHD	26	93.8	13.0	2.5	88.6	99.1	45	100	.892	.49
	All LD	66	97.1	4.6	0.6	96.0	98.2	80	100		
	Affective/PD	29	94.9	9.7	1.8	91.2	98.6	60	100		
	ADHD+LD	20	96.9	5.7	1.3	94.2	99.6	75	100		
	ADHD/LD + Affective/PD	51	95.8	6.9	1.0	93.9	97.8	65	100		
	Cognitive D/O or Neuro	11	96.8	3.4	1.0	94.6	99.1	93	100		
	Total	203	96.0	7.6	0.5	95.0	97.1	45	100		
WMT CNS	All ADHD	26	91.9	12.5	2.4	86.9	97.0	55	100	.611	.69
	All LD	66	95.6	6.5	0.8	94.0	97.2	73	100		
	Affective/PD	29	93.4	11.4	2.1	89.0	97.7	55	100		
	ADHD+LD	20	94.5	6.5	1.4	91.5	97.5	73	100		
	ADHD/LD + Affective/PD	51	93.4	13.7	1.9	89.5	97.2	15	100		
	Cognitive D/O or Neuro	11	94.8	5.5	1.7	91.1	98.5	85	100		
	Total	203	94.1	10.2	0.7	92.7	95.5	15	100		
RDS	All ADHD	35	9.5	1.7	0.3	8.9	10.1	7	14	1.731	.13
	All LD	88	8.8	1.6	0.2	8.4	9.1	5	14		
	Affective/PD	42	9.3	2.3	0.4	8.6	10.0	5	15		
	ADHD+LD	32	8.6	1.8	0.3	7.9	9.3	5	12		
	ADHD/LD + Affective/PD	50	8.7	2.0	0.3	8.1	9.3	5	16		
	Cognitive D/O or Neuro	29	8.3	3.1	0.6	7.1	9.5	0	14		
	Total	276	8.9	2.0	0.1	8.6	9.1	0	16		

Note. CI = Confidence Interval; TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span. ADHD = Attention Deficit/Hyperactivity Disorder; LD = Learning Disorder; PD = Personality Disorder; Cognitive D/O or Neuro = Cognitive Disorder or neurological disorder.

Considering the LPA failed to reveal latent groups, Pass, Fail, and Near Pass groups were created with descriptive statistics parameters and cut-scores for TOMM trial 1, WMT IR, WMT DR, WMT CNS, and RDS. The Pass group was defined as those scoring above the established cut-score for each test. The Near Pass group was defined as those scoring within 1 standard deviation (SD) below the cut-score. The Fail group was defined as those scoring below 1 SD below the cut-score. Table 9 presents the frequencies of these groups for each of the above SVTs. Note again that TOMM trial 2 was not used due to the fact that only four individuals scored below the cut-score.

Table 9: *Frequency Table for Pass, Near Pass, and Fail Groups for SVTs*

	SVT				
Classification	TOMM 1	WMT IR	WMT DR	WMT CNS	RDS
Pass	200	232	234	225	256
Near Pass	14	11	7	15	58
Fail	14	11	12	12	9
Totals	228	254	253	252	324

Note. SVT = Symptom Validity Test; TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span

Using these groupings, each individual was classified as either Pass, Near Pass, or Fail and mean differences in neuropsychological test performance were measured across these classifications. For all such analyses, a Bonferroni Correction was used to control for alpha inflation given the number of comparisons being made. Note, however, that the classifications used here resulted in relatively small sample sizes per group across all SVTs. This attrition is further complicated by the fact that not all individuals were given the same neuropsychological tests. As a result, only the most commonly administered test items could be used to compare groups' neurocognitive functioning. Given that individuals in this study presented for assessment

of academic needs, the most common tests across all individuals were the WAIS (either III or IV; a majority of the total sample were administered version IV) and the Woodcock-Johnson III Tests of Achievement (WJ-III).

The following analyses compare neuropsychological test performance across individuals assigned to Pass, Near Pass, and Fail groups based on TOMM trial 1 performance. These same analyses were not calculated for WMS-III and WMS-IV index scores due to low sample sizes for each group (between zero and four individuals each). The same problem occurred with CVLT data with Near Pass and Fail groups containing one person each. Table 10 details the findings of the WAIS-III and WAIS-IV core indices. There were no statistical differences between groups. Again, the sample sizes are very low for the Near Pass and Fail groups thus diminishing statistical utility and interpretation of these particular findings. Despite non-significant findings, lower mean performances were noted for the Near Pass group on WAIS-III VCI, PRI, and WMI. Lower mean performances were observed for the Fail group on WAIS-III PSI and WAIS-IV VCI, PRI, WMI, and PSI.

Despite there being no significant differences in these cognitive measures across the Pass, Near Pass, and Fail groups, this analysis provides some descriptive data with important implications (see discussion section for additional information). Specifically, a passing performance on the TOMM, or any other SVT for that matter, implies that adequate effort was put forth on the part of the patient and therefore, their data can be interpreted with confidence (i.e., the patient put forth effort, so their data must be valid). Not finding a difference between the Pass, Near Pass, and Fail groups here implies that despite having reduced performance on the TOMM, there was no statistical difference in performance across these IQ indices.

Also, it is interesting that the maximum range on several indices in the Near Pass and Fail group is quite high (WAIS-III PRI has a max of 123[superior range] despite being in the Fail group; the Near Pass group has a maximum of 124 [superior range] on the WAIS-III WMI; the Fail group has a WAIS-IV VCI maximum of 147 [very superior]; and the Fail group has a WAIS-IV PRI maximum of 138[superior]). Of course, this is limited by the fact that the Near Pass and Fail groups had very small sample sizes for the WAIS III and IV. Regardless, these high scores beg the question of the utility of reduced SVT scores on interpretation of cognitive data – can these scores, in the superior range, be interpreted as invalid based on reduced SVT performance? Can a clinician argue that the patient could have done better than superior because SVT performance revealed "sub-optimal" effort?

Table 10: *ANOVA for IQ Testing in Groups Defined by TOMM Trail 1*

						95% CI					
		N	Mean	SD	SE	Lower	Upper	Min	Max	F	*p*
WAIS-III VCI	Pass	40	112.5	13.1	2.1	108.3	116.7	78	136	.272	.76
	Near Pass	3	107.0	9.6	5.6	83.0	131.0	100	118		
	Fail	3	111.0	6.6	3.8	94.7	127.3	105	118		
	Total	46	112.0	12.5	1.8	108.3	115.8	78	136		
WAIS-III PRI	Pass	40	114.4	13.2	2.1	110.2	118.6	78	135	.349	.71
	Near Pass	3	108.0	9.8	5.7	83.5	132.5	97	116		
	Fail	3	114.7	9.1	5.2	92.1	137.2	105	123		
	Total	46	114.0	12.7	1.9	110.2	117.8	78	135		
WAIS-III WMI	Pass	40	102.1	14.6	2.3	97.4	106.7	78	139	.142	.87
	Near Pass	3	98.0	22.5	13.0	42.0	154.0	84	124		
	Fail	3	99.3	6.4	3.7	83.4	115.3	92	104		
	Total	46	101.6	14.5	2.1	97.3	105.9	78	139		
WAIS-III PSI	Pass	40	95.8	12.2	1.9	91.9	99.6	71	122	1.397	.26
	Near Pass	3	88.0	9.6	5.6	64.0	112.0	81	99		
	Fail	3	86.3	4.0	2.3	76.3	96.4	84	91		
	Total	46	94.6	11.9	1.8	91.1	98.2	71	122		
WAIS-IV VCI	Pass	89	109.1	13.8	1.5	106.2	112.0	72	145	.577	.56
	Near Pass	4	105.5	6.6	3.3	95.1	115.9	98	114		
	Fail	7	102.9	36.0	13.6	69.6	136.1	63	147		

	Total	100	108.5	15.9	1.6	105.4	111.7	63	147		
WAIS-IV PRI	Pass	89	105.7	13.7	1.5	102.8	108.6	77	133	1.729	.18
	Near Pass	4	97.5	9.9	5.0	81.7	113.3	84	105		
	Fail	7	96.7	25.1	9.5	73.5	120.0	71	138		
	Total	100	104.7	14.7	1.5	101.8	107.6	71	138		
WAIS-IV WMI	Pass	89	95.5	17.4	1.8	91.8	99.2	14	136	.549	.58
	Near Pass	3	101.7	14.3	8.3	66.2	137.2	86	114		
	Fail	7	89.9	18.1	6.8	73.2	106.6	58	108		
	Total	99	95.3	17.3	1.7	91.8	98.8	14	136		
WAIS-IV PSI	Pass	89	94.9	12.0	1.3	92.3	97.4	71	127	2.619	.08
	Near Pass	3	103.3	5.7	3.3	89.2	117.5	97	108		
	Fail	7	85.1	22.3	8.4	64.5	105.8	50	114		
	Total	99	94.4	13.0	1.3	91.8	97.0	50	127		

Notes. Numerator df was 2 in all cases. WAIS = Wechsler Adult Intelligence Scale; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

When TOMM Trial 1 Pass, Near Pass, and Fail groups were compared on academic achievement testing, multiple comparisons were statistically different (see Table 11). The Fail group scored significantly lower on Oral Language (p = .007), Broad Reading (p = .04), Broad Math (p = .02), and Written Expression (p = .003). However, after Bonferroni Correction, only three comparisons remained significant. The Pass and Fail groups were statistically different from each other on the Oral Language (p = .007), Broad Math (p = .03), and the Written Expression (p = .003) indices.

Most of the indices were not different across groups. Mean performances across the WJ-III indices were in the average range for all SVT classifications. Small sample sizes may have again limited the interpretation of these analyses. Regardless, as noted above in reference to Table 10, it is interesting to note that the maximum performance on some of the WJ-III indices for putatively invalid SVT performance (i.e., the Near Pass and Fail groups), were so high as to question the utility of the SVT. Specifically, the maximum scores for the Broad Math, Brief

Reading, and Brief Writing indices for the SVT Fail group ranged from 132 to 133 (superior range). No research has examined whether clinicians would consider these scores invalid due to reduced SVT performance.

Table 11: *ANOVA for WJ-III in Groups Defined by TOMM Trail 1*

						95% CI					
		N	Mean	SD	SE	Lower	Upper	Min	Max	F	*p*-val
Oral Language index	Pass	76	99.3	11.0	1.3	96.8	101.8	76	125	5.352	.007
	Near Pass	2	91.5	6.4	4.5	34.3	148.7	87	96		
	Fail	5	82.8	16.6	7.4	62.2	103.4	62	105		
	Total	83	98.1	11.9	1.3	95.5	100.7	62	125		
Total Ach. index	Pass	72	98.5	10.5	1.2	96.0	100.9	72	123	2.933	.059
	Near Pass	3	92.0	3.0	1.7	84.5	99.5	89	95		
	Fail	5	87.6	13.0	5.8	71.4	103.8	69	104		
	Total	80	97.5	10.8	1.2	95.1	99.9	69	123		
Broad Reading index	Pass	78	95.9	9.3	1.1	93.8	98.0	71	121	3.336	.040
	Near Pass	3	89.0	2.6	1.5	82.4	95.6	87	92		
	Fail	5	85.8	12.0	5.4	70.9	100.7	68	96		
	Total	86	95.0	9.6	1.0	93.0	97.1	68	121		
Broad Math index	Pass	186	100.2	11.1	0.8	98.6	101.8	65	129	3.979	.020
	Near Pass	12	96.1	7.9	2.3	91.1	101.1	77	110		
	Fail	12	91.3	18.9	5.5	79.2	103.3	58	133		
	Total	210	99.4	11.6	0.8	97.9	101.0	58	133		
Broad Written Lang. index	Pass	117	100.3	11.8	1.1	98.1	102.5	62	125	2.767	.067
	Near Pass	5	99.2	7.5	3.3	89.9	108.5	91	108		
	Fail	7	89.4	15.7	5.9	74.9	103.9	71	108		
	Total	129	99.7	12.0	1.1	97.6	101.8	62	125		
Brief Reading index	Pass	167	97.9	9.2	0.7	96.5	99.3	66	122	0.690	.503
	Near Pass	10	94.8	11.5	3.6	86.6	103.0	68	109		
	Fail	11	95.5	17.2	5.2	84.0	107.1	69	132		
	Total	188	97.6	9.9	0.7	96.1	99.0	66	132		
Brief Math index	Pass	179	97.1	12.2	0.9	95.3	98.9	30	131	2.070	.129
	Near Pass	10	92.4	9.4	3.0	85.6	99.2	71	104		
	Fail	12	90.6	18.5	5.3	78.8	102.3	61	127		
	Total	201	96.5	12.6	0.9	94.7	98.2	30	131		
Math Calc. Skills index	Pass	184	98.6	13.3	1.0	96.7	100.6	58	139	1.439	.240
	Near Pass	14	104.6	35.9	9.6	83.9	125.3	64	220		
	Fail	13	94.2	19.1	5.3	82.7	105.8	54	123		
	Total	211	98.8	16.1	1.1	96.6	100.9	54	220		
Brief Writing index	Pass	171	100.9	12.7	1.0	98.9	102.8	51	130	0.150	.861
	Near Pass	10	99.7	17.6	5.6	87.1	112.3	63	117		
	Fail	11	98.8	19.1	5.8	86.0	111.7	65	133		

	Total	192	100.7	13.3	1.0	98.8	102.6	51	133		
Written Express. index	Pass	122	101.8	11.5	1.0	99.7	103.8	66	131	5.942	.003
	Near Pass	5	97.8	13.8	6.2	80.7	114.9	78	109		
	Fail	7	86.4	13.1	4.9	74.3	98.5	71	104		
	Total	134	100.8	12.1	1.0	98.8	102.9	66	131		
Academic Fluency index	Pass	182	95.0	13.7	1.0	92.9	97.0	17	146	0.436	.647
	Near Pass	14	95.9	15.6	4.2	86.9	104.8	68	118		
	Fail	12	91.2	20.5	5.9	78.2	104.2	49	118		
	Total	208	94.8	14.3	1.0	92.8	96.7	17	146		

Notes. Numerator df was 2 in all cases. Ach. = Achievement; Calc. = Calculation; Express. = Expression.

Unfortunately, due to small sample sizes, Pass, Near Pass, and Fail groups were not compared for the WMT subtests. Reliable Digit Span yielded the highest overall frequency of the SVTs administered which allowed for group comparisons across multiple neuropsychological domains. Results of group differences between WAIS-III and WAIS-IV indices are presented in Table 12. For WAIS-III no RDS failing scores were noted. As such, only Pass and Near Pass groups were compared across the Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed indices. Results indicated that the Pass and Near Pass groups were meaningfully different on working memory with a 17 point difference between these groups ($p < .000$). For the WAIS-IV, very few people failed the RDS (n = 9). Regardless, groups differed significantly on all four WAIS-IV indices with the Fail group performing worse across all indices.

Table 12: *ANOVA for IQ Testing in Groups Defined from RDS*

						95% CI					
		N	Mean	SD	SE	Lower	Upper	Min	Max	F	*p*
WAIS-III VCI	Pass	69	114.1	13.7	1.6	110.8	117.4	78	148	2.801	.098
	Near Pass	14	107.4	12.9	3.4	100.0	114.9	88	134		
	Fail	0	-	-	-	-	-	-	-		
	Total	83	113.0	13.7	1.5	110.0	116.0	78	148		
WAIS-III PRI	Pass	69	115.0	14.0	1.7	111.7	118.4	70	135	2.455	.121
	Near Pass	14	109.0	7.7	2.0	104.6	113.4	97	125		
	Fail	0	-	-	-	-	-	-	-		
	Total	83	114.0	13.3	1.5	111.1	116.9	70	135		
WAIS-III WMI	Pass	69	104.6	12.4	1.5	101.6	107.6	80	139	22.997	.000
	Near Pass	14	87.6	9.9	2.6	81.9	93.3	73	109		
	Fail	0	-	-	-	-	-	-	-		
	Total	83	101.7	13.6	1.5	98.8	104.7	73	139		
WAIS-III PSI	Pass	69	96.7	13.6	1.6	93.4	100.0	71	134	.354	.553
	Near Pass	14	94.4	9.5	2.5	89.0	99.9	81	114		
	Fail	0	-	-	-	-	-	-	-		
	Total	83	96.3	12.9	1.4	93.5	99.1	71	134		
WAIS-IV VCI	Pass	181	112.7	13.5	1.0	110.8	114.7	81	147	15.777	.000
	Near Pass	46	104.8	14.0	2.1	100.6	108.9	63	145		
	Fail	9	90.9	14.8	4.9	79.5	102.3	68	108		
	Total	236	110.3	14.5	0.9	108.5	112.2	63	147		
WAIS-IV PRI	Pass	181	109.8	13.6	1.0	107.8	111.7	73	138	11.606	.000
	Near Pass	46	102.3	11.6	1.7	98.9	105.8	77	129		
	Fail	9	92.8	15.0	5.0	81.2	104.3	71	121		
	Total	236	107.7	13.9	0.9	105.9	109.4	71	138		
WAIS-IV WMI	Pass	178	100.6	17.7	1.3	98.0	103.2	9	136	22.256	.000
	Near Pass	46	88.6	9.6	1.4	85.7	91.4	71	111		
	Fail	9	70.9	7.9	2.6	64.8	77.0	58	80		
	Total	233	97.1	17.6	1.2	94.8	99.3	9	136		
WAIS-IV PSI	Pass	179	97.9	12.6	0.9	96.0	99.7	42	127	6.688	.002
	Near Pass	46	92.4	10.9	1.6	89.2	95.7	68	114		
	Fail	9	85.8	20.0	6.7	70.4	101.1	50	120		
	Total	234	96.3	12.9	0.8	94.7	98.0	42	127		

Notes. Numerator df was 2 in all cases. WAIS = Wechsler Adult Intelligence Scale; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

Again, maximum performances for the Near Pass and Fail groups are quite high despite lower performance on RDS. Also, the Working Memory index is composed of digit span and another measure of mental control. As such, as RDS declines, so does digit span total score, and with it, the Working Memory Index. Accordingly, on this analysis those on the SVT Fail group had the lowest maximum score for WAIS-IV WMI (score of 80).

Table 13 represents the Bonferroni correction for WAIS-IV; most of the comparisons between groups remained significant with the Fail group performing significantly worse in all cases.

Table 13: *Bonferroni Correction for WAIS-IV Across Groups Defined by RDS*

						95% CI	
Dependent Variable			Mean Diff	SE	*p*	Lower	Upper
WAIS-IV (VCI)	Pass	Near Pass	7.97	2.25	0.001	2.5	13.4
		Fail	21.85	4.66	0.000	10.6	33.1
	Near Pass	Fail	13.87	4.97	0.017	1.9	25.9
WAIS-IV (PRI)	Pass	Near Pass	7.41	2.19	0.003	2.1	12.7
		Fail	16.98	4.53	0.001	6.1	27.9
	Near Pass	Fail	9.57	4.84	0.147	-2.1	21.2
WAIS-IV (WMI)	Pass	Near Pass	11.99	2.68	0.000	5.5	18.5
		Fail	29.69	5.53	0.000	16.3	43.0
	Near Pass	Fail	17.7	5.9	0.009	3.5	31.9
WAIS-IV (PSI)	Pass	Near Pass	5.45	2.09	0.029	0.4	10.5
		Fail	12.1	4.31	0.016	1.7	22.5
	Near Pass	Fail	6.66	4.6	0.448	-4.4	17.8

Note. WAIS = Wechsler Adult Intelligence Scale; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

Lastly, the RDS-defined Pass, Fail, and Near Pass groups were compared on academic achievement measures (Table 14). All indices were statistically different with the Pass group having higher mean performance and the Fail group having the lowest mean performance across all indices.

Table 14: *ANOVA for WJ-III Testing Across Groups Defined by RDS*

						95% CI					
		N	Mean	SD	SE	Lower	Upper	Min	Max	F	*p*
Oral Language index	Pass	54	100.1	9.0	1.2	97.7	102.6	81	125	6.214	.003
	Near Pass	23	94.4	13.1	2.7	88.7	100.0	62	114		
	Fail	4	84.3	4.4	2.2	77.2	91.3	80	89		
	Total	81	97.7	10.8	1.2	95.3	100.1	62	125		
Total Ach. Index	Pass	136	101.7	9.4	0.8	100.1	103.3	81	126	15.714	.000
	Near Pass	33	91.4	17.6	3.1	85.1	97.6	8	107		
	Fail	7	86.3	8.2	3.1	78.7	93.9	72	94		
	Total	176	99.2	12.3	0.9	97.3	101.0	8	126		
Broad Reading index	Pass	139	98.8	9.0	0.8	97.3	100.3	81	130	12.528	.000
	Near Pass	34	92.0	8.5	1.5	89.0	95.0	68	107		
	Fail	7	87.1	7.7	2.9	80.0	94.3	71	94		
	Total	180	97.0	9.4	0.7	95.7	98.4	68	130		
Broad Math index	Pass	238	100.0	13.5	0.9	98.3	101.7	9	137	12.515	.000
	Near Pass	56	93.7	11.2	1.5	90.7	96.7	58	122		
	Fail	8	81.3	9.2	3.3	73.5	89.0	69	94		
	Total	302	98.3	13.5	0.8	96.8	99.9	9	137		
Broad Written Lang. index	Pass	160	103.4	10.6	0.8	101.7	105.0	76	132	13.465	.000
	Near Pass	39	96.2	12.4	2.0	92.2	100.2	62	125		
	Fail	8	87.3	12.2	4.3	77.1	97.4	69	100		
	Total	207	101.4	11.7	0.8	99.8	103.0	62	132		
Brief Reading index	Pass	217	99.0	13.4	0.9	97.2	100.8	6	129	7.013	.001
	Near Pass	52	94.3	8.7	1.2	91.9	96.7	75	114		
	Fail	8	85.0	11.3	4.0	75.6	94.4	66	96		
	Total	277	97.7	12.9	0.8	96.2	99.2	6	129		
Brief Math index	Pass	234	99.3	14.0	0.9	97.5	101.1	30	142	7.232	.001
	Near Pass	56	94.9	11.3	1.5	91.9	97.9	61	118		
	Fail	8	83.5	10.4	3.7	74.8	92.2	67	95		
	Total	298	98.1	13.8	0.8	96.5	99.6	30	142		
Math Calc. Skills index	Pass	241	98.9	16.7	1.1	96.7	101.0	54	220	9.898	.000
	Near Pass	58	90.5	17.3	2.3	86.0	95.1	3	122		
	Fail	8	80.3	9.5	3.4	72.3	88.2	67	93		
	Total	307	96.8	17.2	1.0	94.9	98.7	3	220		
Brief Writing	Pass	226	102.6	14.0	0.9	100.8	104.5	10	132	8.760	.000
	Near Pass	54	96.0	11.9	1.6	92.8	99.3	63	122		

index	Fail	8	88.3	12.9	4.5	77.5	99.0	65	102		
	Total	288	101.0	14.0	0.8	99.4	102.6	10	132		
Written Express. Index	Pass	163	104.6	12.2	1.0	102.7	106.5	78	141	9.354	.000
	Near Pass	39	98.4	11.7	1.9	94.6	102.2	66	119		
	Fail	8	89.3	12.2	4.3	79.1	99.4	71	100		
	Total	210	102.8	12.6	0.9	101.1	104.5	66	141		
Academic Fluency index	Pass	237	94.9	13.7	0.9	93.1	96.6	17	143	9.213	.000
	Near Pass	55	87.9	12.2	1.6	84.6	91.2	49	110		
	Fail	8	81.8	6.7	2.4	76.1	87.4	73	94		
	Total	300	93.3	13.6	0.8	91.7	94.8	17	143		

Notes. Numerator df was 2 in all cases. Ach. = Achievement; Calc. = Calculation; Express. = Expression.

Table 15 represents the Bonferroni correction for WJ-III testing. Most of the comparisons between groups remained significant with the Fail group performing significantly worse in all cases.

Table 15: *Bonferroni Correction for WJ-III Across Groups Defined by RDS*

Dependent Variable			Mean Diff.	SE	*p*	95% CI Lower	95% CI Upper
Oral Language index	Pass	Near Pass	5.7	2.54	0.08	-0.49	11.92
		Fail	15.9	5.28	0.01	2.95	28.77
	Near Pass	Fail	10.1	5.52	0.21	-3.36	23.64
Total Ach. index	Pass	Near Pass	10.4	2.20	0.00	5.03	15.69
		Fail	15.4	4.40	0.00	4.79	26.08
	Near Pass	Fail	5.1	4.73	0.85	-6.35	16.51
Broad Reading index	Pass	Near Pass	6.8	1.70	0.00	2.68	10.88
		Fail	11.6	3.43	0.00	3.34	19.93
	Near Pass	Fail	4.9	3.68	0.57	-4.03	13.75
Broad Math index	Pass	Near Pass	6.4	1.93	0.00	1.71	11.01
		Fail	18.8	4.67	0.00	7.51	30.02
	Near Pass	Fail	12.4	4.91	0.04	0.58	24.24

Broad Written Lang. index	Pass	Near Pass	7.1	1.97	0.00	2.40	11.89
		Fail	16.1	3.99	0.00	6.47	25.73
	Near Pass	Fail	9	4.27	0.11	-1.36	19.27
Brief Reading index	Pass	Near Pass	4.7	1.95	0.05	-0.04	9.36
		Fail	14	4.55	0.01	3.01	24.93
	Near Pass	Fail	9.3	4.80	0.16	-2.25	20.87
Brief Math index	Pass	Near Pass	4.4	2.01	0.08	-0.39	9.28
		Fail	15.8	4.85	0.00	4.13	27.51
	Near Pass	Fail	11.4	5.10	0.08	-0.91	23.66
Math Calc. Skills index	Pass	Near Pass	8.3	2.44	0.00	2.48	14.22
		Fail	18.6	6.00	0.01	4.18	33.05
	Near Pass	Fail	10.3	6.29	0.31	-4.88	25.42
Brief Writing index	Pass	Near Pass	6.6	2.06	0.00	1.66	11.60
		Fail	14.4	4.90	0.01	2.59	26.20
	Near Pass	Fail	7.8	5.16	0.40	-4.66	20.20
Written Express. index	Pass	Near Pass	6.2	2.16	0.01	1.00	11.43
		Fail	15.3	4.39	0.00	4.73	25.91
	Near Pass	Fail	9.1	4.70	0.16	-2.25	20.46
Academic Fluency index	Pass	Near Pass	7	1.99	0.00	2.17	11.75
		Fail	13.1	4.77	0.02	1.64	24.63
	Near Pass	Fail	6.2	5.03	0.66	-5.92	18.28

Notes. Numerator df was 2 in all cases. Ach. = Achievement; Calc. = Calculation; Express. = Expression.

In summary, results of the current analysis provide valuable data regarding the prevalence of SVT failure in an academic accommodation seeking population. Data also describe the differential performance on different SVTs with those having a diagnosis performing worse on TOMM trial 1 and RDS. Results indicate that most people pass SVTs, despite diagnostic status. The Latent Profile Analysis results indicate that no class solution (*k*) was statistically better than another. Thus, no single class solution was identified as a reasonable fit for the data.

Those with a diagnosis of any kind were no more likely to be classified as having failed an SVT than those without a diagnosis and there were no differences in SVT outcomes across the different diagnostic classifications. A few differences were noted in cognitive and academic achievement testing scores for the Pass, Near Pass, and Fail groups. An interesting finding here was that despite having a Near Pass or Fail classification on any given SVT, maximum performances on some tests were in the superior to very superior range, begging the question of how fruitful it is to say that reduced SVT performance invalidates all other test performances.

In the next section, implications of these findings will be discussed in the context of other published works. Strengths and limitations will also be discussed in detail.

Discussion

Prevalence of SVT Failure

The current study sought to measure the failure rate of various SVTs in academic accommodation seeking college students. This is an important aspect of the study because of the variable base rates of SVT failure reported in other studies. Base rates range from 8% in medical cases to 30% in disability cases (Mittenberg et al., 2002) and as high as 40% in clinical cases (Larrabee, 2003). In assessment of college students seeking academic accommodations, several studies likewise suggest that base rates of malingering are considerably high. One study found that 22% of college students feigned ADHD symptoms and another study found that as many as 50% of college students fake ADHD symptoms (Sullivan et al., 2007; Marshall et al., 2010). A more recent study suggests that in asymptomatic undergraduates not seeking secondary gain, suboptimal effort, as measured by three standard SVTs, was as high as 55% (An et al., 2013).

In the current study, SVT failure rates ranged from 2 to 21%. Importantly, many of the students evaluated in the current study presented with questions of ADHD, LD, or a combination

of the two, in addition to other psychiatric and cognitive impairments. Accordingly, it is important to compare base rates found in the current study to those reported among similar clinical populations with similar demographics. The failure rate for the WMT in the current study varied by WMT subtest and ranged from 8 to 11 percent across some 250 administrations. This finding is noteworthy as it was derived from a larger sample size as compared to other similar studies. The 8-11 percent failure on the WMT reported here is drastically different than other studies with similar populations. For instance, Sullivan et al. (2007) provided data on 66 students presenting or ADHD and/or LD assessment at a university clinic. Demographics for the students in the Sullivan et al. study were stratified by diagnosis but were similar to those presented in this current study (Table 1). Authors used the WMT with standard cut-scores to define SVT failure. Of all individuals assessed, 22.4 percent were found to have suboptimal performance, a figure twice as high as that reported in the current study. A 31 percent failure rate on the WMT was also reported by Suhr et al. (2008) with a similar sample of 85 college students referred for ADHD evaluations. It is unclear what could account for the difference in the 8-11 percent WMT failure rate noted in the present investigation and the 22-31 percent noted on other studies with ostensibly similar populations and referral questions. One possible explanation is that the current study was composed of a broader patient population whereas the data from the Sullivan et al. and Suhr et al. studies were based on ADHD and/or LD alone. In the current study, 21 percent of the total sample received no diagnosis after assessment, 12.9 percent were diagnosed with affective or personality conditions, and 6.1 percent were diagnosed with Cognitive Disorder NOS or some other neurological condition (e.g., traumatic brain injury). It is possible that other such conditions more readily pass the WMT thus lowering the overall failure rate in this sample. Indeed, other studies support the conclusion that individuals with depression

typically pass the WMT (Rholing, Green, Allen, & Iverson, 2002) and additional studies indicate that individuals with genuine psychological conditions generally pass SVTs (Merten & Merckelbach, 2013). However, it is noteworthy that WMT performance did not differ across diagnostic groups in the current study (see Table 8). As such, the difference in base rates does not appear to be a factor of diagnosis status. In fact, in the current investigation, those with a diagnosis performed better on the WMT IR, WMT DR, and the WMT CNS than those without a diagnosis, although $p > .05$ in all cases.

In the present investigation, out of 231 administrations of TOMM Trial 2, only four individuals fell below the traditional cut-score. This is in line with recent findings from Silk-Eglit et al. (2014) of an average failure rate of 2.26 percent across multiple SVTs in a non-clinical sample of undergraduate college students. However, in the same study by Silk-Eglit and colleagues, no individual fell below the cut score for the TOMM or RDS and failure performances were only found on the WMT (1% of the sample) and the Victoria Symptom Validity Test (3% of the sample). Although the failure rate in TOMM Trail 2 is low in the current study, it is important to note that this is not drastically higher than the 0% found in the non-clinical sample reported in the Silk-Eglit study. The current study, on the other hand, found that out of 324 administrations of the RDS, 21% fell below the cut score. This is a considerably higher fail rate compared to the 0% in the Silk-Eglit study and is likely due to the fact that individuals in the current study were being assessed for cognitive impairments, which may be related to lower performance on digit span testing and in turn, lower RDS scores. Indeed, the mean RDS scores in the current study were significantly different between those with any diagnosis compared to those with no diagnosis. Similarly, the mean RDS performance in the current study was 9.07 (SD = 2.01; range = 4-16) whereas the mean performance reported in the

Silk-Eglit study was 10.18 (SD = 1.89; range = 7-15). As such, the higher fail rate and lower mean performance reported in the current study may be due to the fact that the current study comprised a clinical sample.

In fact, the different failure rates among SVTs in the current study could be related to the fact that each SVT engages a different cognitive network (Allen et al., 2007; Browndyke et al., 2008) and relates to variability in cognitive functioning seen in clinical samples relative to healthy or non-clinical samples. The idea that SVTs employ some level of cognitive engagement, and that testing could be more cognitively taxing on certain individuals, has been suggested in recent work by Bigler (2014). In his argument, Bigler (2014) points out that SVTs engage Top-Down cognitive processing and that individuals with well-documented cognitive impairments that affect attentional and motivational brain systems may demonstrate suboptimal SVT performance.

Experimental studies also support this conclusion. For example, Brooks (2012) examined performance on the Victoria Symptom Validity Test (VSVT) in a sample of 100 pediatric patients with various neurological disorders (traumatic brain injury, stroke, epilepsy, hydrocephalus, and other conditions) with no secondary gain. This researcher also measured whether VSVT performance was related to cognitive functioning on formal testing. The VSVT is divided into easy items and difficult items and test results are typically presented as “valid”, “questionable”, or “invalid.” Results indicate that although no individual outright failed (i.e., “invalid” performance), 5 percent fell in the “questionable” performance range across all items. Notably, the author reported that 97 percent had valid performance on easy items while only 84 percent had valid performance on difficult items and suggested that this could in fact be based on

intellectual functioning rather than effort per se. Indeed, the author found that age and intellectual functioning accounted for 46 percent of the variance on difficult item performance.

In a study of 120 epilepsy patients without external incentive to malinger, VSVT performance predicted Full Scale IQ, and performance on measures of attention and memory (Loring et al. 2005). Similarly, Macciocchi, Seel, Alderson, and Godsall (2006) found that although almost individuals with acute brain injury pass the VSVT, injury severity was predictive of difficult item response latencies (slower performance for worse injury as defined by initial Glasgow Coma Scale score) and that response latencies on difficult items and overall VSVT performance was related to worse neuropsychological test performance. Armistead-Jehle, Gervais, and Green (2012) found that worse SVT performance was correlated with subjective memory complaints in adults with no apparent secondary gain. Keary et al. (2013) found that SVT outcome in neurologic patients was related to memory and intellectual ability. Additional research with older adults with different levels of dementia but no apparent secondary gain found that mild dementia patients performed better on SVTs than those with moderate to severe dementia and that SVT outcomes were related to cognitive function, especially learning capacity (Rudman, Oyebode, Jones, & Bentham, 2011).

All of these studies support the notion that diminished cognitive capacity can play a role in SVT outcome. Unfortunately, the role of cognitive load and capacity is often underappreciated in SVT research (Bigler, 2014). Leighton et al., (2014) reported that recognition memory is typically thought of as an “automatic process” (p. 880) but when cognitive load is increased during the learning phase of SVTs, performance declines. In the current study, cognitive demand may or may not be different across the different SVTs, but cognitive capacity most definitely varies across diagnostic groups. Individuals in this study diagnosed with some type of cognitive

or psychological condition seem to perform worse than those without a diagnosis, at least on TOMM trial 1 and RDS.

In addition, non-neurologic factors such as secondary gain or negative self-expectancy and diagnosis threat may also explain the difference between the current findings and the failure rates reported by Silk-Eglit et al. (2014). Specifically, in the current study, it could be argued that the individuals being assessed may do more poorly in order to secure academic accommodations or to qualify for additional accommodations and that individuals expect to do poorly because they are being assessed for impairment (see below for additional discussion on the role of secondary gain and diagnosis threat).

Latent Groups and a "Gray Zone" in SVT Outcomes

The current study used Latent Profile Analysis (LPA) to determine whether a "Near-Pass" group exists in a given distribution of SVT scores. The idea that three possible latent distributions may exist in a distribution of SVT scores was postulated by Bigler (2012) where he suggested that there may be a group that clearly pass, a group that outright fails (at or below chance performance) and a group that hovers around the cut-score (i.e., the "Near-Pass" group). However, the results of the LPA failed to identify any specific number of latent classes as a reasonable fit for the data. This finding suggests that, at least in the data set used in this study, there is no latent "Near-Pass" group that fall in a gray zone that can be meaningfully separated from other latent groups. Although the present study failed to identify clear latent groups, additional analyses indicated that individuals with a diagnosis performed more poorly on SVTs (See Table 4), specifically on trial 1 of the TOMM and Reliable Digit Span (RDS).

The finding that individuals with a diagnosis performed more poorly on SVTs is important considering the evidence presented above, and the argument proposed by Bigler

(2014), that neurocognitive impairment can result in suboptimal but legitimate performance on SVTs. This finding is supported by ample evidence from other published works – i.e., individuals that fall just over the cut-score may be unique, may have true neurocognitive impairments, and may not in fact be exhibiting poor effort. For example, in a small case study, Willis et al. (2011) demonstrated that even mild cognitive impairments can influence the results of SVTs. Another study found that asymptomatic individuals with various neurological conditions and patients with mild Alzheimer's disease frequently fall below SVT cut-scores (Merten, et al., 2007). In a similar study, 9.7-percent of Mild Cognitive Impairment (MCI) and 21.4-percent of moderate to severe dementia patients fell just below the cut-score of TOMM trial 2 (Walter, et al., 2014). Patients with Huntington's Disease are more likely to fail SVTs as their symptoms increase (Sieck, et al., 2013). Neuropsychiatric patients are also likely to fall below cut-scores of SVTs (Gorissen, et al., 2005). There are multiple studies that support this same theme - patient's with well-documented conditions commonly fail SVTs (Greve, et al., 2008; Howe & Loring, 2009; Keary et al., 2013; Eichstaedt et al., 2014), and that these conditions typically result in scores just below the cut-score (Biger, 2014).

This accumulation of research supports findings in the current study - individuals diagnosed with some type of condition (ADHD, LD, etc.) demonstrated lower performance on SVTs. However, there are multiple weaknesses in this argument. First, although the current study demonstrated that individuals with a diagnosis performed more poorly on TOMM trial 1 and RDS, the mean performance for these SVTs was still above the cut-score typically associated with the test. In addition, when comparing the expected and observed frequencies of SVT failure (i.e., 1 SD below the cut score) and diagnosis status (diagnosis v. no diagnosis) results of the chi-square analysis indicate that for the most part, frequency rates of SVT failure were no more than

expected for individuals with a diagnosis compared to those without. Next, as is true of most symptom validity studies, the role of secondary gain needs to be explored. Secondary gain is defined as the presence of factors that provide external incentive to feign impairment or exaggerate symptoms (Fishbain, Rosomoff, Cutler, & Rosomoff, 1995; Dersh, Polatin, Leeman, & Gatchel, 2004). In the current study, academic accommodations may act as considerable secondary gain for students. Students are often faced with a rigorous and competitive academic environment. There are students who may hope to gain entrance to advanced graduate or medical programs; students may be motivated to gain any advantage they can, including having accommodations to make their schooling more manageable. For instance, those who acquire some form of accommodations may be given extended time on tests, prolonged due dates for assignments, private test environments, alternative courses, and shortened workloads. In addition, in some universities, student health centers will not prescribe medications for ADHD until a student has been screened by the university's academic accommodation office (M. Brooks, personal communication, November 15, 2012). As such, obtaining a prescription for a stimulant medication may also act as a secondary gain factor for students (Sullivan et al., 2007; Musso & Gouvier, 2012).

In the current study, students referred to the university's accommodation center were undoubtedly referred with the understanding that such accommodations might be attainable and that having accommodations may help them perform better in school. However, secondary gain may be less of an issue in the present investigation considering the relatively low SVT fail rates compared to other published studies of academic accommodation seeking students. As noted above, high base rates of SVT failure are reported in multiple studies. In studies with academic populations seeking accommodations, rates of poor performance on the WMT appear to range

from 14-percent in one study (Harrison & Edwards, 2010) to 25-48% in another study (Sullivan et al., 2007). In the current study, failure on the WMT ranged from 8-11%, depending on the WMT subtest. There is also evidence that secondary gain factors are not always predictive of SVT performance. In an examination of 436 referrals to a state psychiatric hospital with some individuals being committed civilly, and some committed to the forensic unit, authors postulated that forensic patients would have more secondary gain and would, therefore, fail SVTs more often. Out of a total of 81 occurrences of failure on the TOMM, 48 were related to secondary gain (forensic criminal changes, fines, etc.) and 31 had no apparent secondary gain. The authors of this study suggested that failure on the TOMM without secondary gain might have been related to cognitive and psychiatric variables (Marcopulos et al., 2014). In addition, although 48 individuals failed the TOMM, 133 individuals with putative secondary gain passed. Accordingly, secondary gain does not always predict SVT outcomes.

In the current study, individuals were referred to a student accessibilities center with questions of ADHD and/or LD, in addition to other cognitive and psychological factors. Given their understanding of the nature of the assessment, individuals in this study could have taken on a sick-role. The adoption of a sick-role based on expectation is referred to as “diagnosis threat” (Suhr & Gunstad, 2002). In one study, Suhr and Gunstad (2002) examined the possible impact of diagnosis threat on individuals with mild traumatic brain injury (mTBI). Some individuals were told that they were selected because of their injury history and that the nature of their injury was associated with cognitive deficits. Others were assigned to a neutral group and were told that they were going to complete tests of cognitive functioning. Those in the diagnosis threat group performed more poorly than the neutral group on measures of memory and intelligence. In a follow-up study, Suhr and Gunstad (2005) further explored diagnosis threat while attempting to

control for other factors known to influence test performance (i.e., depression, anxiety, and effort). Again, with a sample of mTBI patients, some were assigned to a neutral condition and some to a diagnosis threat condition. As with their 2002 study, these researchers found worse cognitive test performances in the diagnosis threat condition. It was interesting to note, however, that depression, anxiety, and effort failed to mediate the effects of diagnosis threat on cognitive test performance. Also, in a more recent study with a sample more similar to that of the current study, Trontel, Hall, Ashendorf, and O'Connor (2013) examined the impact of diagnosis threat on academic self-efficacy (belief in one's own academic abilities) among 54 college students with self-reported mTBI. Using similar conditions as reported by Suhr and Gunstad (2002 & 2005), these authors found that diagnosis threat was related to lower academic self-efficacy.

In summary, the current study demonstrated that individuals with some type of diagnosis performed more poorly on two SVTs (TOMM trial 1 and RDS), although this could have been due to non-neurological factors such as secondary gain or diagnosis threat. Findings also suggest that this could have been due to the fact that individuals with a diagnosis have reduced performance due to attenuated Top-Down cognitive proficiency. This is supported by the fact that very few people performed at, or below chance on any SVT and by the fact that the mean scores on SVTs, regardless of diagnosis status, were above standard cut-scores for the test. The next section will discuss findings related to specific diagnostic classifications.

Cognitive Functioning and SVT Outcomes

The current investigation proposed that individuals who have SVT performances just below the cut-score may in fact have legitimate neurocognitive or psychiatric conditions and due to such diagnoses, would have worse neuropsychological test performance relative to individuals who pass SVTs and to those who perform at or below chance. This was examined by first

operationalizing a Pass group, a Near Pass group, and a Fail group. As noted above, the Pass group was defined as those scoring above the established cut-score for each test. The Near Pass group was defined as those scoring within 1 standard deviation (SD) below the cut-score. The Fail group was defined as those scoring below 1 SD below the cut-score. These groups were compared on multiple cognitive and academic achievement measures. Almost without exception, the Fail group performed more poorly than the Near Pass and Pass groups. In fact, there was a linear relationship between scores on cognitive/academic achievement testing and SVT outcome – as scores on SVTs declined, so did neuropsychological and achievement scores.

This dose-response relationship (Bigler, 2015) and the finding that reduced SVT performance predicts lower neuropsychological test performance is supported by multiple independent investigations. For instance, in an archival study of 53 individuals undergoing medico-legal evaluations, worse performance on various SVTs was associated with lower scores on all but one test of the Halstead–Reitan Neuropsychological Battery (Silk-Eglit, Stenclik, Miele, Lynch, & McCaffrey, 2013). Suchy, Chelune, Franchow, and Thorgusen (2012) examined effort and neuropsychological functioning among 507 individuals independently diagnosed with Multiple Sclerosis without secondary gain. The authors sought to examine whether SVT performance was related to cognitive functioning and to assess whether confronting patients with suboptimal performance improves SVT and cognitive performance. These researchers demonstrated that SVT performance was related to cognitive functioning, supporting the notion of a dose-response relationship. Further, they also demonstrated that after confronting those with suboptimal SVT performance, both SVT and cognitive test performance improved.

In another study that is more analogous to the current investigation, among 144 college students being assessed for academic accommodations, approximately 15-percent failed either

the WMT or the Medical Symptom Validity Test (MSVT) and individuals with failing SVT scores performed more poorly on measures of memory, IQ, and non-verbal processing speed (Harrison & Edwards, 2010). Sullivan et al. (2007) also demonstrated that failure on the WMT predicts worse cognitive functioning. They examined 66 individuals assessed for ADHD, LD or both and used the WMT as their stand-alone measure of performance validity. In the sample, 22% were found to have suboptimal performance on the WMT. Among individuals being evaluated for ADHD, WMT performance correlated positively with Full Scale IQ and performance on the California Verbal Learning Test-2nd edition (i.e., lower, or worse WMT performance was predictive of lower cognitive test scores). Suhr et al. (2008 & 2011) also demonstrated that reduced WMT performance among individuals referred for ADHD evaluations was predictive of reduced neuropsychological test performance. Numerous studies have found that healthy individuals who were asked to feign ADHD symptoms performed worse on measures of attention and processing speed relative to controls (for complete review, see Musso & Gouvier, 2012). In fact, individuals asked to fake ADHD symptoms perform worse on measures of attention and processing speed compared to individuals with true ADHD (Quinn, 2003; Harrison, Edwards, & Parker, 2007) suggesting again that intentional poor effort affects and invalidates all test measures. In one study, authors examined SVT and neuropsychological test performance among 87 patients with unequivocal acquired brain injury (traumatic brain injury, stroke, tumor, anoxic injury, & electrocution). In their sample, 21.8 percent fell below the established cut-score of the TOMM and performance on the TOMM was predictive of poorer outcomes in neuropsychological testing (Locke, Smigielski, Powell, & Stevens, 2008).

This overwhelming evidence that lower SVT performance is associated with lower cognitive testing performance might again be related to secondary gain and feigned

neurocognitive impairment – i.e., given that individuals choose to present themselves as impaired, lower performances are seen across all measures. This is most likely true in cases where individuals perform at or below chance on SVTs – they are likely making a conscious choice to perform poorly (Bigler, 2014). However, this conclusion is more difficult to make when individuals have well-documented cognitive or psychological conditions and when SVT performance falls just below the cut-score (e.g., 44 on the TOMM). An important finding of the current investigation is that very few individuals in a large sample size score at or below chance on any SVT – mean performances and standard deviations for each SVT are noted in Table 2. Given the relatively low base rates of SVT failure in this study compared to other published works, and that those with a diagnosis performed more poorly on SVTs it is possible that true neurocognitive and psychiatric impairment could account for reduced neuropsychological test performance. As brain systems are the neurologic foundation of motivation and drive, it stands to reason that neurological and neuropsychiatric conditions that affect motivational states could have at least some impact on SVT performance, even if only by a few points (Bigler, 2014), and in turn, cognitive test performance.

Accordingly, it is important to ask whether suboptimal performance on SVTs completely invalidates cognitive test performance in the presence of objective evidence of neurological or psychiatric impairment. In his 2014 article, Bigler presents multiple cases of individuals with traumatic brain injury in which injury was confirmed via neuro-imaging but performance on SVTs was suboptimal (sometimes with only one point below cut-scores). Bigler also illustrates that motivational changes are common in brain injury and other neurological and psychiatric conditions (e.g., schizophrenia) and that such factors likely play a role in test performance.

In one investigation, authors examined effort performance and cognitive functioning in 44 individuals with moderate to severe TBI. Patients were initially found to have passing scores on an embedded effort measure on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). They then sought to examine whether scores on embedded measures of effort on the RBANS predicted cognitive functioning on other cognitive measures. These authors found that the effort index of the RBANS was predictive of other cognitive functions even after controlling for education and injury severity. They concluded that effort should be considered on a continuum rather than a dichotomous variable (Lippa, Agbayani, Hawes, Jokic, & Caroselli, 2014).

In summary, results of this study are consistent with other published works – reduced SVT performance could be related to genuine impairments, there is a dose-response relationship between SVTs and neuropsychological test outcomes, and the dose-response relationship implies that SVT performance lies on a continuum. The following section will deal more with the clinical implications of these points.

Implications of Failed SVTs on Clinical Data

A neuropsychologist's ability to make sound clinical interpretation of test data is dependent on an accurate measurement of a patient's cognitive abilities. If a patient puts forth reduced effort, for any reason, and the clinician is unaware of this poor effort, the chances of making a false-positive statement about the patient's condition increases. As such, measuring performance validity is prudent in clinical and forensic settings. However, little work has been done to examine whether passing SVTs equates to full effort or just adequate effort and if differences in adequate effort and full effort have an impact on neuropsychological test results.

As discussed in the previous section, numerous studies document a dose-response relationship between SVT outcome and neuropsychological test findings – as SVT performance declines, so do neuropsychological test scores (Bigler, 2014). Indeed, the current study is largely in line with these previously published works. As such, SVTs appear to have some predictive value for neuropsychology test performance, although this predictive value may be unidirectional (discussed below). The association between SVT and cognitive data is irrelevant when performances on forced-choice SVTs are at or near chance (≤ 50% performance). A dose-response relationship is likely to be a more clinically meaningful phenomenon in the so called "gray zone" (i.e., those that perform just below the cut-score on any given SVT; Bigler, 2013 & 2014), especially given the above discussion that genuine impairments can results in slight declines in SVT outcomes. Advocates of strict cut-scores on SVTs maintain that measurement error and confidence intervals are not critical for SVTs because the tests are intentionally designed to be easy and can therefore be dichotomized as pass or fail (Green & Flaro, 2003; Sweet, 1999; Carone, 2014, see also Bigler, 2014). In other words, if full effort is put forth by a patient, a maximum score is expected and achieving a maximum score is putatively easy. Falling below a cut-score is therefore thought to reflect a level of effort low enough as to warrant other tests of cognition invalid or uninterpretable. However, there are a few concerns with this position.

First, one concern is that the use of cut-scores ignores the psychometric phenomenon of measurement error. There is also a concern with the utility of predicting effort on neuropsychological test data with a measure that has a low ceiling. A ceiling effect occurs because a majority of tested individuals perform at or near the maximum score on a given measure. When a majority of individuals perform at the ceiling, a clinician losses the ability to

discriminate among patients at the upper end of the scale. In other words, it may be possible that two individuals with a maximum TOMM score of 50 could actually be putting forth different levels of effort. This limits the upper end of the dose-response relationship discussed above – as SVT performance declines, so do performances on other measures but as SVT performance increases, the ceiling effect reduces the predictive power at the upper end of other measures. Limited variability on one measure reduces a test's ability to predict another variable.

In fact, there is research support for the idea that initially passing an SVT does not necessarily indicate that maximum effort has been achieved. As noted above, Suchy et al. (2012) confronted MS patients about their questionable performance on SVTs. When patients were confronted, two-thirds improved their performance to a valid range on the Victoria Symptom Validity Test (VSVT), which also resulted in improved performance on the cognitive measures. Another study examined the effects on monetary incentive on prospective memory tests among children with moderate to severe TBI. Although there was less of an impact on the severe TBI patients, higher incentive did improve performance among the moderate TBI children (McCauley et al., 2011). An important point about the Suchy et al. and the McCauley et al. papers is that the patient's diagnosis was not in question and there was no apparent motivation to do poorly.

Next, these studies support the idea that effort is not a binary variable but rather lies on a continuum. In fact, methodological studies also indicate that effort, as measured by SVTs is a continuous variable, not categorical (Walters et al., 2008; Walters, Berry, Rogers, Payne, & Granacher, 2009). In one study, authors used three different nonredundant taxometric procedures to examine the latent structure of three different SVTs – the VSVT, the TOMM, and the Letter Memory Test (LMT). Taxometric procedures included three well-known means of determining if

a variable is dimensional or taxonic – the *mean above minus below a cut* (MAMBAC), *maximum covariance* (MAXCOV), and *latent-mode factor analysis* (L-Mode).Their sample consisted of 527 compensation seeking adults referred for neuropsychological evaluation. Overall, the authors showed that across all measures, performance was on a continuum and suggested that effort should not be thought of as a dichotomous construct (Walters et al., 2009). Using similar methods (MAMBAC and MAXCOV) this same group examined the latent structure of SVTs for psychopathology among 1211 adults. Results were similar with authors concluding that feigned psychopathology occurs on a continuum rather than a taxon (Walters et al, 2008). In fact, Walters et al. (2009) concluded that there may be need for an intermediate classification for those that do not clearly pass or clearly fail SVTs.

Finally, no study to date has examined the interpretation of variable cognitive data observed in the presence of sub-optimal SVT performance. In other words, low SVT performance is often used to invalidate low cognitive test performance. However, it is unclear whether average performances on cognitive testing are invalid when SVT performance is low or if low performance on cognitive data is observed in only one cognitive domain (e.g., low attention and working memory but average IQ and executive abilities). The current study raises several key limitations in describing all cognitive test performance as invalid when SVT performance is suboptimal. For instance, as discussed above, it is interesting that the maximum range on several indices in the Near Pass and Fail group is quite high. This was true for the WAIS-III, WAIS-IV, and WJ-III. Although the sample sizes in some instances were low, these high scores beg the question of the utility of reduced SVT scores on interpretation of cognitive data – can above average scores be interpreted as invalid based on reduced SVT performance?

Can a clinician argue that the patient could have done better than superior because SVT performance revealed “sub-optimal” effort?

To further illustrate this point, Table 16 was complied based on cognitive data from Tables 10 and 12 above. For simplicity, only above average performances based on standard Wechsler Classifications were extracted (Wechsler Classifications describe data as average, high average, superior, etc.; Wechsler, 2008).

Table 16: *Above Average Performances with Suboptimal SVT Performance*

Reference Table	SVT Classification	Test Variable	n	Max Score	Wechsler Classification
Table 10	Near Pass	WAIS-III VCI	3	118	High Average
	Fail	WAIS-III VCI	3	118	High Average
	Near Pass	WAIS-III PRI	3	116	High Average
	Fail	WAIS-III PRI	3	123	Superior
	Near Pass	WAIS-III WMI	3	124	Superior
	Near Pass	WAIS-IV VCI	4	114	High Average
	Fail	WAIS-IV VCI	7	147	Very Superior
	Fail	WAIS-IV PRI	7	138	Very Superior
	Near Pass	WAIS-IV PSI	3	114	High Average
	Fail	WAIS-IV WMI	7	114	High Average
Table 12	Near Pass	WAIS-III VCI	14	134	Very Superior
	Near Pass	WAIS-III PRI	14	125	Superior
	Near Pass	WAIS-IV VCI	46	145	Very Superior
	Near Pass	WAIS-IV PRI	46	129	Superior
	Fail	WAIS-IV PRI	9	121	Superior
	Near Pass	WAIS-IV WMI	46	111	High Average
	Near Pass	WAIS-IV PSI	46	114	High Average
	Fail	WAIS-IV PSI	9	120	Superior

Note. WAIS = Wechsler Adult Intelligence Scale; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index.

The implications here are important because of previous research that suggests that the specificity of some SVTs is as high as 100% (Institute of Medicine, 2015). In other words, if failure on an SVT (performance below an established cut-score) unequivocally means poor

effort, then even the performances in Table 16 would be considered invalid. Bigler (2012 & 2014) suggests that such an argument would undoubtedly increase false-positive classifications on the part of the clinician. Clinically speaking, data from Table 16 imply that poor SVT performance cannot possibly invalidate all cognitive data. In the recent Institute of Medicine document cited above, "…even in the context of PVT failure, performances that are in the average range can be interpreted as reflecting ability that is in the average range or above, though such performances may represent an underestimate of actual level of ability." (IOM, 2015; section 5, p. 17). Clearly more research needs to be done to improve performance validity testing and minimize classification errors.

Strengths & Limitations

This study, by its very nature, was quasi-experimental and as such, brings with it methodological limitations. First, there was obviously no experimental manipulation of an independent variable as this study was conducted *ex post facto*. Indeed, archival studies such as this lack control over any variable. Considering quasi-independent variables were used, no causal statements can be made in the current investigation.

The current study has a much larger sample size than a majority of SVT-related research articles. With a total sample of ~540 individuals, this study was amenable to multiple analyses. In other studies with similar populations, it is common for authors to utilize one or two SVTs. Although one study concluded that the use of multiple SVTs can drastically increase false-positive rates of poor effort (Berthelson et al., 2013), the current study is unique in that it analyzed several SVTs with total administrations ranging from 228 for TOMM trial 1 to 324 for RDS. However, despite these large samples, this study was archival in nature and based on a clinical sample. As such, different testing procedures were used for each individual. Of the total

sample, not all individuals were administered SVTs and of those that were administered an SVT, only a small subset scored below traditional cut-scores (see Table 2). Also, of those who were administered an SVT of any type, not all were administered every neuropsychological test. For instance, since this is archival data, some earlier participants were administered the WAIS-III or WMS-III while more recent assessments included WAIS-IV and WMS-IV, while others may not have had these tests at all. Accordingly, usable sample sizes for any given comparison are considerably smaller than expected and multiple cognitive variables were simply lost to attrition of data.

Another strength of this study is the nature in which individuals were diagnosed (or not diagnosed) following formal assessment. For each student assessed in BYU's Accessibility Center, strict DSM-IV criteria were used in diagnosis. However, in addition to diagnostic criteria, group consensus (i.e., doctoral level clinical psychologists and a clinical neuropsychologist) was employed to determine the diagnosis of each individual. Thus, the diagnostic classifications presented in this study may be more accurate than other studies where assessment was based on single-clinician judgment or where undergraduates are asked to fake symptoms in malingering analogue studies. Despite the strength that comes from clear diagnostic classification per group, a significant limitation in this study comes from the difficulty in combining classifications into larger groups. As noted in Figure 2 and Table 5 above, there were nuanced differences between sub-classifications which resulted in multiple conditions being combined. Although this allowed for simplified statistical exploration and increased sample size per classification, it could certainly be argued that significant heterogeneity exists within any given classification. Some likely have more heterogeneity than others. For instance, all individuals with one diagnosis of any type of learning disability (LD) were combined into an LD

group. The same was true of those with a subtype of an ADHD diagnosis. All individuals with an LD and ADHD diagnosis were combined into a single ADHD+LD group. Thus, ADHD and/or LD groups were likely composed of individuals with similar complaints. In fact, other studies similar this one also combine ADHD and/or LD diagnoses (Sullivan et al., 2007; Suhr et al., 2008). There were, of course, many individuals with more than one diagnosis. All individuals with a type of LD, ADHD, and an axis I or II affective or personality disorder classification were combined into a single ADHD/LD + Affective/PD group. The last category was more difficult to define and represents a significant limitation to this study – there were many individuals with neurological conditions (e.g., seizure disorders, TBI, & brain tumor) and individuals diagnosed simply as "Cognitive Disorder NOS". It was assumed that those with Cognitive Disorder NOS must not have met criteria for ADHD, LD, or any other type of axis I or II condition. As such, all individuals with a neurological disorder and all individuals with a diagnosis of Cognitive Disorder NOS were combined into a single group. However, this was likely the most heterogeneous group. In addition, there were 12 individuals that did not seem to fit any of the above classifications (e.g., Asperger syndrome, cerebral palsy, Developmental Coordination Disorder, various visual impairments, etc.). These individuals were excluded from the descriptive statistics above.

Conclusions & Future Directions

Despite lack of support for the original hypotheses, results of the current study provide important data regarding the prevalence of SVT failure in an academic accommodation seeking population. This study hypothesized that the "Near Pass" group postulated by Bigler (2012) may exist and that these individuals will meet criteria for either a neuropsychological or psychiatric condition more often than other latent groups. Although this hypothesis was not supported in the

current investigation, there was other evidence that diagnostic status influences performance on SVTs. Results indicated that a majority of the sample pass SVTs, despite possible influence of non-neurological factors such as secondary gain and diagnosis threat. In addition, the current study examined whether neuropsychological test performance differs between groups defined as Pass, Near Pass, and Fail status with the hypothesis that the Near Pass group would perform worse given the assumption that this group would most likely consist of those receiving a neuropsychological or psychiatric diagnosis. Again, this was not supported in the current paper, although a few differences were noted in cognitive and academic achievement with the Pass group most consistently outperforming the Near Pass and Fail group. A linear or dose-response relationship was observed with lower SVTs performance being associated with lower mean performance on cognitive and academic testing. An interesting finding here was that despite having a Near Pass or Fail classification on any given SVT, maximum performances on some tests were in the superior to very superior range, begging the question of how fruitful it is to say that reduced SVT performance invalidates all other test performances.

Overall, findings support the idea that future research needs to consider SVT performance as being on a continuum rather than as a dichotomous variable. In addition, future research needs to clarify the role of non-neurological factors and how they influence SVT performance. In their review of factors that affect invalid neuropsychological test performance among head injured individuals Iverson and Binder (2000) state the following, “Magnification of symptoms or suboptimal effort on neuropsychological tests can have several independent or related underlying causes. Therefore, detecting nonneurological symptom reporting or atypical effort within the context of a neuropsychological evaluation does not automatically indicate that the individual is malingering.” (p. 830) and “The well-informed clinician will seek to identify all

variables that may affect symptom reporting or neuropsychological test performance and be careful not to over- or under interpret evidence of negative response bias." (p. 853).

The Institute of Medicine (2015), in a recent statement regarding the use of psychological tests for disability determination, proposed that objective medical evidence was required in all disability evaluations. They further indicated that in order for psychological testing to be considered objective, it must be validated by means of standard PVTs (performance validity tests; heretofore referred to as SVTs). In fact, they further recommend that a statement about the test validity be provided in all disability evaluation reports. However, the IOM recognizes that slight declines in SVT performance do not necessarily constitute invalid test performance, "A lack of validity on performance validity testing alone is insufficient grounds for denying a disability claim. In such cases, additional information is required to assess the applicant's allegation of disability." (IOM; p. S-8).

These statements are echoes of previously published consensus statements about the use of SVTs. Bush et al. (2005) suggest that formal tests are not the only way to measure effort and that the assessment of effort also includes clinical observation. In addition, Bush et al. point out that SVTs that fall slightly below cut-off may not imply malingering and that additional indicators are needed first. In the AACN consensus statement cited above, we read the following: "Whether using a multivariable composite or a single test, neuropsychologists should not rely on single, fixed cut scores. Neuropsychologists appreciate and consider a range of cut scores and associated diagnostic test statistics in choosing the cut score to be applied to a specific case. The decision-making process occurs in different contexts, such that the relative costs of false positive and false negative errors will not be constant across situations." Authors continue, "To assist clinicians in this decision-making process, investigators, journal editors, and test publishers are

strongly encouraged to provide a broad range of cut scores with their respective diagnostic test statistics (e.g., sensitivities, specificities, and likelihood ratios)." (Heilbronner et al., 2009, p. 27).

The difficulty of correctly classifying those that put forth poor effort is further discussed in detail by Bigler (2013 & 2014) – how can a clinician state that test findings are invalid and that a patient is feigning impairment when so many documented cases exist of low SVT performance among genuinely impaired individuals without secondary gain? Without objective evidence that a patient is healthy (i.e., no brain impairment), clinicians run the risk of classifying an impaired individual as malingering. To remedy this problem, we first need to change our classification scheme of effort from dichotomous to continuous and recognize that individuals in the "gray zone" may represent a unique clinical population. The studies by Walters et al. (2008 & 2009) noted above support the idea that further work needs to be done on a Near Pass group. These authors reported the following, "The dimensional results observed in the present study indicate that a dichotomous decision may not be appropriate and that it may be useful to create an indeterminate category for marginal scores; this indeterminate category could be based on the standard error of measurement for the purpose of reducing misclassifications." (p. 591).

Future work needs to follow the recommendations of the above consensus statements and seek additional information that could account for poor performance on SVTs. For example, studies have not classified differential performance based on demographics and medical history (Bigler, in press). A majority of SVTs were developed with artificial groups asked to feign impairment. Research has not examined the differential impact of diagnosis, or other demographic factors on SVT outcome. One particular SVT, The Dot Counting Test, is one of the few that stratifies cut-scores by different diagnoses (depression, schizophrenia, head injury,

stroke, learning disability, mild & dementia; Boone, Lu, & Herzberg, 2002). Similar adjustable cut-scores for other commonly used SVTs may minimize misclassification of poor effort.

Bigler, (in press) notes that no study to date has examined the use of neuroimaging in determination of cut-scores for any given SVT. Specifically, the author suggests that future work examine the role of cortical atrophy and hippocampal volume and whether lesion burden or damage to specific brain networks has an impact on SVT performance. He further suggests that structural and functional brain imaging may be useful in distinguishing if low SVT performance is related to true neurological impairment versus psychiatric factors, such as conversion or somatoform disorders.

In conclusion, the results of the present study lend support to previous research that reduced performance on SVTs may be related to neuropsychological or psychiatric factors and that reduced SVT performance is related to poorer outcomes on cognitive and academic achievement testing. The presence of an intermediate or Near Pass group needs further investigation. Future research needs to address the use of cut-scores as such binary taxonomies likely results in misclassification of individuals. Future research needs to establish how non-neurological factors influence SVT outcomes and the field should move towards more empirical means of measuring performance validity.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Allen, M. D., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2007). Functional neuroimaging evidence for high cognitive effort on the word memory test in the absence of external incentives. *Brain Injury, 21*(13-14), 1425-1428.

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 27*(8), 849-857.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Armistead-Jehle, P., Gervais, R.O., & Green, P. (2012). Memory complaints inventory and symptom validity test performance in a clinical sample. *Archives of Clinical Neuropsychology, 27*(7), 725-34.

Batt, K., Shores, E. A., & Chekaluk, E. (2008). The effect of distraction on the word memory test and test of memory malingering performance in patients with a severe brain injury. *Journal of the International Neuropsychological Society, 14*(6), 1074-1080.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*(7-8), 909-916.

Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*(1), 19-45.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland digit recognition test. *Journal of Clinical and Experimental Neuropsychology, 15*(2), 170-182.

Bigler, E.D. (2011). Effort – What is it, how should it be measured? *Journal of the International Neuropsychological Society, 17*, 751-752.

Bigler, E.D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 632-640.

Bigler, E.D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*, 13-14.

Bigler, E.D. (2015). *Use of symptom validity tests and performance validity tests in disability determinations.* Institute of Medicine. Washington, DC: The National Academies Press.

Bigler, E.D. (in press). Neuroimaging as a biomarker in symptom validity and performance validity testing. *Brain Imaging and Behavior, x*(x), x-x.

Boone, K. B., Lu, P., & Herzberg, D. S. (2002). *The Dot Counting Test manual*. Los Angeles, CA: Western Psychological Services.

Brooks, B.L. (2012). Victoria Symptom Validity Test performance in children and adolescents with neurological disorders. *Archives of Clinical Neuropsychology, 27*(8), 858-868.

Browndyke, J.N., Paskavitz, J., Sweet, L.H., Cohen, R.A., Tucker, K.A., Welsh-Bohmer, K.A., Burke, J.R., & Schmechel, D.E. (2008). Neuroanatomical correlates of malingered memory impairment: Event-related fMRI of deception on a recognition memory task. *Brain Injury, 22*(6), 481–489.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., . . . Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology, 20*(4), 419-426.

Carone, D. A. (2014). Young child with severe brain volume loss easily passes the word memory test and medical symptom validity test: Implications for mild TBI. *The Clinical Neuropsychologist, 28*(1), 146-162.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 28*(7), 633-639.

Cizek, G. (1993). Reconsidering standards and criteria. *Journal of Educational Measurement*, *30*(2), 93-106.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (2000). *California verbal learning test manual*, adult version (2nd ed.). San Antonio, TX: Psychological Corporation.

Dersh, J., Polatin, P.B., Leeman, G., & Gatchel, R.J. (2004). The management of secondary gain and loss in medicolegal settings: Strengths and weaknesses. *Journal of Occupational Rehabilitation, 14*(4), 267-79.

Dwyer, C. (1996). Cut scores and testing: Statistics, judgment, truth, and error. *Psychological Assessment, 8*(4), 360-362.

Eichstaedt, K. E., Clifton, W. E., Vale, F. L., Benbadis, S. R., Bozorg, A. M., Rodgers-Neame, N. T., & Schoenberg, M. R. (2014). Sensitivity of green's word memory test genuine memory impairment profile to temporal pathology: A study in patients with temporal lobe epilepsy. *The Clinical Neuropsychologist, 28*(6), 941-953.

Fishbain, D.A., Rosomoff, H.L., Cutler, R.B., & Rosomoff, R.S. (1995). Secondary gain concept: A review of the scientific evidence. *Clinical Journal of Pain, 11*(1), 6-21.

Gavett, B.E., O'Bryant, S.E., Fisher, J.M., & McCaffrey, R.J. (2005). Hit rates of adequate performance based on the test of memory malingering (TOMM) Trial 1. *Applied Neuropsychology, 12*(1), 1-4.

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word memory test performance in amnesic patients with hippocampal damage. *Neuropsychology, 23*(4), 529-534.

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

Gouvier, W. D., Lees-Haley, P., & Hammer, J. H. (2003). The neuropsychological examination in the problem of detecting malingering in the forensic arena: Costs and benefits. In G. P. Prigatano & N. H. Pliskin (Eds.), *Clinical neuropsychology and cost outcomes research: A beginning* (pp. 405–424). New York: Psychology Press.

Graham, J.R. (2011). *MMPI-2: Assessing personality and psychopathology,* (3rd ed.). New York, NY: Oxford University Press, Inc.

Green, P. (2003). *Green's word memory test for Windows: User's manual*. Edmonton, Alberta, Green's Publishing Inc.

Green, P., & Flaro, L. (2003). Word memory test performance in children. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 9*(3), 189-207.

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22*(5), 896-918.

Harrison, A.G., & Edwards, M.J. (2010). Symptom exaggeration in post-secondary students: Preliminary base rates in a Canadian sample. *Applied Neuropsychology, 17*, 135-143.

Harrison, A. G., Edwards, M. J., & Parker, K. C. (2007). Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology, 22*(5), 577-588.

Harrison, A. G., Flaro, L., & Armstrong, I. (2014). Rates of effort test failure in children with ADHD: An exploratory study. *Applied Neuropsychology: Child, 25*, 1-14.

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

Howe, L. L., & Loring, D. W. (2009). Classification accuracy and predictive ability of the medical symptom validity test's dementia profile and general memory impairment profile. *The Clinical Neuropsychologist, 23*(2), 329-342.

Institute of Medicine. (2015). *Psychological testing in the service of disability determination.* Washington, DC: The National Academies Press.

Iverson, G.L., & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 15*(2), 829-58.

Jasinski, L. J., Harp, J. P., Berry, D. T., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *The Clinical Neuropsychologist, 25*(8), 1415-1428.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425-461.

Keary, T. A., Frazier, T. W., Belzile, C. J., Chapin, J. S., Naugle, R. I., Najm, I. M., & Busch, R. M. (2013). Working memory and intelligence are associated with victoria symptom validity test hard item performance in patients with intractable epilepsy. *Journal of the International Neuropsychological Society : JINS, 19*(3), 314-323.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the medical symptom validity test. *The Clinical Neuropsychologist, 24*(5), 860-872.

Lanza, S.T., & Rhoades, B.L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*(2), 157-168.

Larrabee, G.J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*(3), 410-425.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625-630.

Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the word memory test. *Brain Injury, 24*(2), 89-99.

Leighton, A., Weinborn, M., & Maybery, M. (2014). Bridging the gap between neurocognitive processing theory and performance validity assessment among the cognitively impaired: A review and methodological approach. *Journal of the International Neuropsychological Society : JINS, 20*(9), 873-886.

Lippa, S.M., Agbayani, K.A., Hawes, S., Jokic, E., & Caroselli, J.S. (2014). Effort in acute traumatic brain injury: Considering more than pass/fail. *Rehabilitation Psychology, 59*(3), 306-12.

Lo., Y., Mendell, N.R., & Rubin, D.B. (2001) Testing the number of components in a normal mixture. *Biometrika*, *88*(3):767-778.

Locke, D.E., Smigielski, J.S., Powell, M.R., & Stevens, S.R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation, 23*(3), 273-281.

Loring, D. W., Lee, G. P., & Meador, K. J. (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., Marino, S. E., Drane, D. L., Parfitt, D., Finney, G. R., & Meador, K. J. (2011). Lorazepam effects on word memory test performance: A randomized, double-blind, placebo-controlled, crossover trial. *The Clinical Neuropsychologist, 25*(5), 799-811.

Macciocchi, S.N., Seel, R.T., Alderson, A., & Godsall, R. (2006). Victoria symptom validity test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology, 21*(5):395-404.

Marcopulos, B.A., Caillouet, B.A., Bailey, C.M., Tussey, C., Kent, J., & Frederick, R., (2014). Clinical decision making in response to performance validity test failure in a psychiatric setting. *The Clinical Neuropsychologist, 28*(4), 633-652.

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., & Barker, J. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*(7), 1204-1237.

McCauley, S.R., Pedroza, C., Chapman, S.B., Cook, L.G., Vasquez, A.C., & Levin, H.S. (2011). Monetary incentive effects on event-based prospective memory three months after

traumatic brain injury in children. *Journal of Clinical and Experimental Neuropsychology, 33*(6), 639–646

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 308-318.

Merten, T., & Merckelbach, H. (2013). symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*(2), 112-137.

Mittenberg, W., Patton, C., Canyock, E.M., & Condit, D.C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8), 1094-1102.

Musso, M. W., & Gouvier, W. D. (2014). "Why is this so hard?" A review of detection of malingered ADHD in college students. *Journal of Attention Disorders, 18*(3), 186-201.

Muthen, B.O., Muthen, L.C. (2011) *Mplus Version 6.01*. Los Angeles: Muthen and Muthen.

Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: An Interdisciplinary Journal, 14*(4), 535-569.

O'Bryant, S.E., Engel, L.R., Kleiner, J.S., Vasterling, J.J., & Black, F.W. (2007).Test of memory malingering (TOMM) trial 1 as a screening measure for insufficient effort. *The Clinical Neuropsychologist, 21*(3), 511-521.

O'Bryant, S.E., Gavett, B.E., McCaffrey, R.J., O'Jile, J.R., Huerkamp, J.K., Smitherman, T.A., & Humphreys, J.D. (2008). Clinical utility of trial 1 of the test of memory malingering (TOMM). *Applied Neuropsychology, 15*(2), 113-116.

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*(4), 379-395.

Rholing, M.L., Green, P., Allen, L.M., & Iverson, G.L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology, 17*(3), 205-222.

Rudman, N., Oyebode, J.R., Jones, C.A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health. 15*(1), 47-57.

Schreoder, R.W., Twumasi-Ankran, P., Baade, L.E., & Marshall, P.S. (2012). Reliable digit span: Systematic review and cross-validation study. *Assessment, 19*(1), 21-30.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Sclove, S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3):333–43.

Sieck, B. C., Smith, M. M., Duff, K., Paulsen, J. S., & Beglinger, L. J. (2013). Symptom validity test performance in the Huntington disease clinic. *Archives of Clinical Neuropsychology, 28*(2), 135-143.

Silk-Eglit, G. M., Stenclik, J. H., Gavett, B. E., Adams, J. W., Lynch, J. K., & Mccaffrey, R. J. (2014). Base rate of performance invalidity among non-clinical undergraduate research participants. *Archives of Clinical Neuropsychology, 29*(5), 415-421.

Silk-Eglit, G.M., Stenclik, J.H., Miele, A.S., Lynch, J.K., & McCaffrey, R.J. (2013). The degree of conation on neuropsychological tests does not account for performance invalidity among litigants. *Archives of Clinical Neuropsychology, 28*(3), 213-221.

Slick, D.J., Sherman, E., & Iverson, G. (1999) Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545-561.

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26*(8), 774-789.

Suchy, Y., Chelune, G., Franchow, E.I., & Thorgusen, S.R. (2012). Confronting patients about insufficient effort: The impact on subsequent symptom validity and memory performance. *The Clinical Neuropsychologist, 26*(8), 1296-1311.

Suhr, J.A., & Gunstad, J. (2002). "Diagnosis threat": The effect of negative expectations on cognitive performance. *Journal of Clinical and Experimental Neuropsychology*, *24*(4), 448– 457.

Suhr, J. A., & Gunstad, J. (2005). Further exploration of the effect of "diagnosis threat" on cognitive performance in individuals with mild head injury. *Journal of the International Neuropsychological Society, 11*(1)*,* 23–29.

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology, 23*(5), 521-530.

Suhr, J. A., Sullivan, B. K., & Rodriguez, J. L. (2011). The relationship of noncredible performance to continuous performance test scores in adults referred for attention-deficits/hyperactivity disorder evaluation. *Archives of Clinical Neurospychology, 26*(1), 1-7.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*(3), 189-207.

Sweet, J.J. (1999) Malingering: differential diagnosis. In Sweet JJ (Ed) *Forensic neuropsychology: Fundamentals and practice*. Swets & Zeitlinger, New York, pp. 255–285.

Trontel, H.G., Hall, S., Ashendorf, L., & O'Connor, M.K. (2013). Impact of diagnosis threat on academic self-efficacy in mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 35*(9), 960-970.

Van Dyke, S.A., Millis, S.R., Axelrod, B.N., & Hanks, R.A. (2013). Assessing effort: differentiating performance and symptom validity. *The Clinical Neuropsychologist, 27*(8), 1234-1246.

Walter, J., Morris, J., Swier-Vosnos, A., & Pliskin, N. (2014). Effects of severity of dementia on a symptom validity measure. *The Clinical Neuropsychologist, 28*(7), 1197-1208.

Walters, G. D., Berry, D. T., Rogers, R., Payne, J. W., & Granacher, R. P., Jr. (2009). Feigned neurocognitive deficit: Taxon or dimension? *Journal of Clinical and Experimental Neuropsychology, 31*(5)*,* 584–593.

Walters, G. D., Rogers, R., Berry, D. T., Miller, H. A., Duncan, S. A., McCusker, P. J., Payne, J.W., & Granacher, R. P. (2008). Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment, 20*(3)*,* 238–247.

Wechsler, D. (2002). *WAIS–III/WMS–III technical manual, updated.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition: Technical and interpretive manual.* San Antonio, TX: Pearson.

Willis, P. F., Farrer, T. J., & Bigler, E. D. (2011). Are effort measures sensitive to cognitive impairment? *Military Medicine, 176*(12), 1426-1431.

Young, J.C., & Gross, A.M. (2011). Detection of response bias and noncredible performance in adult attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology, 26*(3),165-175.

Zieky, M.J. (2001). So much has changed: How the setting of cutscores has evolved since the 1980s. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives.* (pp. 19-51). Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Appendix A
Institutional Review Board for Human Subjects Approval Letter

Institutional Review Board
for Human Subjects

Brigham Young University
A-285 ASB Provo, Utah 84602
(801) 422-3841 / Fax: (801) 422-0620

May 29, 2015

Thomas Farrer
346 W. 400 North #3
Provo, Ut 84601

Re: E 130237
Passing or Failing of Symptom Validity Tests: Neuropsychological Assessment of "Near-Pass" Patients

Dear Thomas Farrer

This is to inform you Brigham Young University's IRB has renewed its approval of the above noted research study.

The approval period is from 5-29-2015 to 6-26-2016. Your study number is E130237. Please be sure to reference either this number and/or the study title in any correspondence with the IRB.

All conditions for continued approval during the prior approval period remain in effect. These include, but are not necessarily limited to the following requirements:

All protocol amendments and changes to approved research must be submitted to the IRB and not be implemented until approved by the IRB.

Sincerely,

Allen Parcell, PhD., Chair
Sandee M.P. Munoz, Administrator
Institutional Review Board for Human Subjects

Appendix B – Consent Form

University Accessibility Center (UAC)
Treatment and Evaluation Policy

Please Read the Following information carefully

Mission Statement: University Accessibility Center (UAC) is committed to the pursuit of excellence and the realization of human potential. UAC seeks to provide all students with disabilities equal access to the educational opportunities and to eliminate barriers which impede participation in academic pursuits at BYU.

Who we are: The full-time staff of UAC is made up of dedicated professionals who have training in psychology, counseling, education, American Sign Language interpreting, assistive technology, and related fields.

Confidentiality: It is the policy of UAC that all information discussed with a counselor is confidential. Written, telephone, or personal inquiries about you will not be acknowledged. No information about you or your accommodations can be released to anyone other than you without your written permission. An exception to this policy may arise in situations of reported child abuse or where there is a clear and present danger to yourself or others. Please discuss this with your counselor if you have any questions.

Appointments: It is best to set up an appointment with a counselor in advance. Due to scheduling limitations walk-in appointments may not be available. Please be prompt for an appointment and notify us of cancellations or the need to reschedule.

Ethical and Professional Standards: The UAC counselors and graduate students are required to uphold the highest level of professional standards and ethics. If you have any questions or concerns about your services through UAC, please discuss these issues with your counselor or the Director of UAC.

Resolving Disagreements: In the event that you feel you are not being treated fairly, disagree with the conclusions of your counselor regarding a diagnosis and functional limitations, or feel that you are not being offered reasonable accommodations for your disability you can contact the BYU Equal Opportunity Office at 422-5895.

Test Results: It is the policy of UAC to present test results in such a way as to be understandable. UAC will not release test results directly to a student without an explanation. Typically raw test results will not be released to a student. Your test results can be sent directly to another professional only with your written release.

Research: UAC is involved in ongoing research in an attempt to better understand various types of disabilities. The results of learning disability tests may be used ***without your identifying data*** to better understand how BYU can meet the needs of students with learning disabilities. If your test results are used for research purposes, there will be no way to identify the test results as yours.

Services for Students with Disabilities
Consent for Treatment, Accommodative Services, and/or Evaluation

Name:_____ BYU ID:_____

Date of Birth:_____Today's Date:_____

Local Address:_____

_____

Local Phone Number:_____Work Phone Number:_____

I have read the information regarding policies and practices for University Accessibility Center (UAC) and consent to treatment, counseling, accommodative services, and/or evaluation.

Client Signature:_____ Date:_____

Witness:_____ Date:_____

g:\forms\LD\LD-Intake 2009.doc

Appendix C

Post-hoc analysis of difference between diagnostic classifications across SVTs

Variable	(I)	(J)	(I-J)	SD	*p*	95% I Lower	95% I Upper
TOMM Trial 1	All ADHD	All LD	-1.28	0.98	0.78	-4.1	1.5
		Affective/PD	0.65	1.07	0.99	-2.4	3.7
		ADHD+LD	-0.73	1.24	0.99	-4.3	2.9
		ADHD/LD + Affective/PD	-0.83	1.05	0.97	-3.8	2.2
		Cognitive D/O or Neuro	0.46	1.12	1.00	-2.8	3.7
	All LD	Affective/PD	1.94	0.93	0.30	-0.8	4.6
		ADHD+LD	0.55	1.12	1.00	-2.7	3.8
		ADHD/LD + Affective/PD	0.45	0.90	1.00	-2.2	3.1
		Cognitive D/O or Neuro	1.74	0.99	0.49	-1.1	4.6
	Affective/PD	ADHD+LD	-1.38	1.21	0.86	-4.9	2.1
		ADHD/LD + Affective/PD	-1.48	1.00	0.68	-4.4	1.4
		Cognitive D/O or Neuro	-0.19	1.08	1.00	-3.3	2.9
	ADHD+LD	ADHD/LD + Affective/PD	-0.10	1.18	1.00	-3.5	3.3
		Cognitive D/O or Neuro	1.19	1.25	0.93	-2.4	4.8
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	1.29	1.06	0.83	-1.8	4.4
TOMM Trial 2	All ADHD	All LD	-1.23	0.54	0.22	-2.8	0.3
		Affective/PD	0.17	0.59	1.00	-1.5	1.9
		ADHD+LD	-1.04	0.67	0.62	-3.0	0.9
		ADHD/LD + Affective/PD	-0.95	0.58	0.57	-2.6	0.7
		Cognitive D/O or Neuro	-0.52	0.62	0.96	-2.3	1.3
	All LD	Affective/PD	1.40	0.52	0.09	-0.1	2.9
		ADHD+LD	0.19	0.60	1.00	-1.6	1.9
		ADHD/LD + Affective/PD	0.28	0.50	0.99	-1.2	1.7
		Cognitive D/O or Neuro	0.71	0.56	0.80	-0.9	2.3

	Affective/PD	ADHD+LD	-1.21	0.65	0.43	-3.1	0.7
		ADHD/LD + Affective/PD	-1.12	0.56	0.34	-2.7	0.5
		Cognitive D/O or Neuro	-0.69	0.61	0.86	-2.4	1.1
	ADHD+LD	ADHD/LD + Affective/PD	0.09	0.63	1.00	-1.7	1.9
		Cognitive D/O or Neuro	0.52	0.68	0.97	-1.4	2.5
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	0.43	0.59	0.98	-1.3	2.1
WMT IR	All ADHD	All LD	-3.81	2.10	0.46	-9.8	2.2
		Affective/PD	-0.78	2.41	1.00	-7.7	6.2
		ADHD+LD	-3.22	2.69	0.84	-11.0	4.5
		ADHD/LD + Affective/PD	-1.64	2.18	0.98	-7.9	4.6
		Cognitive D/O or Neuro	-1.57	3.26	1.00	-11.0	7.8
	All LD	Affective/PD	3.03	1.97	0.64	-2.7	8.7
		ADHD+LD	0.59	2.31	1.00	-6.1	7.2
		ADHD/LD + Affective/PD	2.17	1.69	0.79	-2.7	7.0
		Cognitive D/O or Neuro	2.23	2.95	0.97	-6.3	10.7
	Affective/PD	ADHD+LD	-2.44	2.60	0.94	-9.9	5.0
		ADHD/LD + Affective/PD	-0.86	2.06	1.00	-6.8	5.1
		Cognitive D/O or Neuro	-0.79	3.18	1.00	-9.9	8.4
	ADHD+LD	ADHD/LD + Affective/PD	1.58	2.39	0.99	-5.3	8.5
		Cognitive D/O or Neuro	1.65	3.40	1.00	-8.1	11.4
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	0.07	3.01	1.00	-8.6	8.7
WMT DR	All ADHD	All LD	-3.24	1.76	0.44	-8.3	1.8
		Affective/PD	-1.07	2.05	1.00	-7.0	4.8
		ADHD+LD	-3.03	2.25	0.76	-9.5	3.5
		ADHD/LD + Affective/PD	-1.99	1.83	0.89	-7.3	3.3
		Cognitive D/O or Neuro	-2.97	2.73	0.88	-10.8	4.9

	All LD	Affective/PD	2.17	1.69	0.79	-2.7	7.0
		ADHD+LD	0.21	1.94	1.00	-5.4	5.8
		ADHD/LD + Affective/PD	1.25	1.41	0.95	-2.8	5.3
		Cognitive D/O or Neuro	0.27	2.47	1.00	-6.8	7.4
	Affective/PD	ADHD+LD	-1.96	2.20	0.95	-8.3	4.4
		ADHD/LD + Affective/PD	-0.92	1.76	1.00	-6.0	4.2
		Cognitive D/O or Neuro	-1.90	2.68	0.98	-9.6	5.8
	ADHD+LD	ADHD/LD + Affective/PD	1.04	2.00	1.00	-4.7	6.8
		Cognitive D/O or Neuro	0.06	2.85	1.00	-8.1	8.3
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	-0.98	2.52	1.00	-8.2	6.3
WMT CNS	All ADHD	All LD	-3.68	2.38	0.63	-10.5	3.2
		Affective/PD	-1.44	2.78	1.00	-9.4	6.6
		ADHD+LD	-2.58	3.06	0.96	-11.4	6.2
		ADHD/LD + Affective/PD	-1.46	2.48	0.99	-8.6	5.7
		Cognitive D/O or Neuro	-2.85	3.70	0.97	-13.5	7.8
	All LD	Affective/PD	2.24	2.29	0.92	-4.4	8.8
		ADHD+LD	1.11	2.62	1.00	-6.5	8.7
		ADHD/LD + Affective/PD	2.22	1.92	0.86	-3.3	7.7
		Cognitive D/O or Neuro	0.83	3.35	1.00	-8.8	10.5
	Affective/PD	ADHD+LD	-1.14	2.99	1.00	-9.7	7.5
		ADHD/LD + Affective/PD	-0.02	2.39	1.00	-6.9	6.9
		Cognitive D/O or Neuro	-1.41	3.64	1.00	-11.9	9.1
	ADHD+LD	ADHD/LD + Affective/PD	1.12	2.71	1.00	-6.7	8.9
		Cognitive D/O or Neuro	-0.27	3.86	1.00	-11.4	10.8
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	-1.39	3.42	1.00	-11.2	8.5

RDS	All ADHD	All LD	0.71	0.40	0.49	-0.5	1.9
		Affective/PD	0.20	0.46	1.00	-1.1	1.5
		ADHD+LD	0.89	0.49	0.47	-0.5	2.3
		ADHD/LD + Affective/PD	0.79	0.45	0.49	-0.5	2.1
		Cognitive D/O or Neuro	1.21	0.51	0.17	-0.3	2.7
	All LD	Affective/PD	-0.51	0.38	0.76	-1.6	0.6
		ADHD+LD	0.18	0.42	1.00	-1.0	1.4
		ADHD/LD + Affective/PD	0.07	0.36	1.00	-1.0	1.1
		Cognitive D/O or Neuro	0.50	0.43	0.86	-0.8	1.7
	Affective/PD	ADHD+LD	0.69	0.47	0.69	-0.7	2.1
		ADHD/LD + Affective/PD	0.59	0.42	0.74	-0.6	1.8
		Cognitive D/O or Neuro	1.01	0.49	0.31	-0.4	2.4
	ADHD+LD	ADHD/LD + Affective/PD	-0.11	0.46	1.00	-1.4	1.2
		Cognitive D/O or Neuro	0.32	0.52	0.99	-1.2	1.8
	ADHD/LD + Affective/PD	Cognitive D/O or Neuro	0.42	0.47	0.95	-0.9	1.8

Note. TOMM = Test of Memory Malingering; WMT = Word Memory Test; IR = Immediate Recall; DR = Delayed Recall; CNS = Consistency; RDS = Reliable Digit Span. ADHD = Attention Deficit/Hyperactivity Disorder; LD = Learning Disorder; PD = Personality Disorder; Cognitive D/O or Neuro = Cognitive Disorder or neurological disorder.