Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2016-06-01

The Use of Embedded and Stand-Alone Measures of Effort in Predicting Academic Ability in College Students

Danita Renee Williams
Brigham Young University

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

 Part of the Counseling Psychology Commons

BYU ScholarsArchive Citation

Williams, Danita Renee, "The Use of Embedded and Stand-Alone Measures of Effort in Predicting Academic Ability in College Students" (2016). *All Theses and Dissertations*. 6434.
https://scholarsarchive.byu.edu/etd/6434

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.

The Use of Embedded and Stand-Alone Measures of Effort in Predicting Academic Ability in College Students

Danita Renee Williams

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

Doctor of Philosophy

Derek Griner, Chair
Michael Brooks
Lane Fischer
Ellie L. Young
Mark E. Beecher

Department of Counseling Psychology and Special Education

Brigham Young University

June 2016

Copyright © 2016 Danita Renee Williams

All Rights Reserved

ABSTRACT

The Use of Embedded and Stand-Alone Measures of Effort in Predicting Academic Ability in College Students

Danita Renee Williams
Department of Counseling Psychology and Special Education, BYU
Doctor of Philosophy

Detection of sub-optimal effort is a critical element of all psychological assessment procedures. Failure to consider the validity of the client's performance and symptom reporting may result in inaccurate conclusions about the degree of impairment. Because the American with Disabilities Act requires colleges to provide accommodations for students with documented disabilities, providing resources for students feigning impairment may ultimately drain university resources intended to help those students with disabilities. This study sought to examine the relationship between two different types of measures of effort and variables related to academic ability. De-identified archival data was gathered from the University Accessibility Center (UAC) at Brigham Young University (BYU) which provided psychological assessments for accommodation seeking students ($N = 602$) for a reduced fee. Measures used to detect sub-optimal effort included the Test of Memory Malingering (TOMM), Word Memory Test (WMT), Validity Indicator Profile (VIP), California Verbal Learning Test-Second Edition (CVLT-II), Reliable Digit Span (RDS), and the Integrated Visual and Auditory Continuous Performance Advanced Edition (IVA-AE). Measures indicating academic ability included select subtests from the Woodcock Johnson Test of Achievement Third Edition (WJ-III). Additionally, Matrix Reasoning of the Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV) was included as a cognitive measure of nonverbal IQ. Two point biserial correlations were conducted. Results indicated that the nonverbal portion of the VIP had a significant relationship with writing fluency. The TOMM also had a significant relationship with writing fluency. Additionally, results demonstrated that Reliable Digit Span had a significant relationship with Academic Fluency, Writing Fluency, Letter Word Identification, and Math Fluency. Data suggests that university disability service offices may wish to include the RDS, TOMM, and VIP in their considerations of effort.

Keywords: accommodation, sub-optimal effort, college student population, TOMM, WMT, VIP

ACKNOWLEDGMENTS

I would like to thank my committee, especially my chair, Dr. Derek Griner, for their assistance, support, and patience as I worked toward finishing my college career. I would also like to thank my internship training director, Dr. Brian Thomson, for his unrelenting support, kindness, and push toward the finish line. Lastly, I would like to thank my family and friends, especially my parents, Kim and Linda Williams, for showing me the truth that is Philippians 4:13.

TABLE OF CONTENTS

LIST OF TABLES

DESCRIPTION OF DISSERTATION STRUCTURE

This dissertation, *The Use of Embedded and Stand-Alone Measures of Effort in Predicting Sub-Optimal Effort in College Students*, is written in a hybrid format. The hybrid format brings together traditional thesis requirements with journal publication formats. This facilitates submitting the dissertation for publication in research and education journals.

The preliminary pages of the thesis reflect requirements for submission to the university. The dissertation is presented as a journal article, and conforms to length and style requirements.

The literature review is included in Appendix A.

This dissertation format contains two reference lists. The first reference list contains references included in the journal-ready article. The second list includes all citations used in the Appendix entitled “Review of the Literature.”

Introduction

Detection of suboptimal effort is a critical element of any psychological assessment (Bush et al., 2005; Larrabee, 2012). Without tests of effort, many clinicians consider neuropsychological batteries to be incomplete (Bauer, O'Bryant, Lynch, McCaffery, & Fisher, 2007). In order to express confidence in the test scores, diagnoses, and treatment recommendations, there must be confidence that the tests were both administered and taken consistent with developer guidelines, meaning they were given to the population intended and under the same or similar circumstances as those on which the test was normed. (Bush et al., 2005). Green, Rohling, Lees-Haley, and Allen (2001) completed a study in which they discovered that the results of their statistical analyses of gathered data changed significantly when data was restricted to include only the participants that passed tests of effort. Consequently, failure to consider the validity of the client's performance and symptom reporting may result in inaccurate conclusions about the degree of impairment (Larrabee, 2012). Because financial resources are often limited in facilities that perform psychological testing, test administrators have an interest in providing efficient and accurate testing results.

Malingering and Sub-Optimal Effort

Malingering is primarily defined as the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs (American Psychiatric Association, 2013). The Diagnostic and Statistical Manual – Fifth Edition (DSM-5) delineates several criteria that must be met in order for the diagnosis of malingering to apply to behavior. The DSM-5 also advises that combinations of any of the following create suspicion for malingering:

1. Medicolegal context of presentation (e.g., the individual is referred by an attorney to the clinician for examination, or the individual self-refers while litigation or criminal charges are pending).
2. Marked discrepancy between the individual's claimed stress or disability and the objective findings and observations.
3. Lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen.
4. The presence of Antisocial Personality Disorder.

A diagnosis of malingering may be distinguished from the use of the terms "poor effort" or "sub-optimal effort." For a clinician to make a diagnosis of malingering, the clinician likely considers several sources of information, including: diagnostic/clinical interviews, reviews of medical records, reviews of any supplemental reports, academic records and history, etc. In circumstances where a formal diagnosis of malingering may not be warranted, clinicians may use the term "sub-optimal effort" to describe a pattern of poor effort. The term *symptom validity tests* may be used to describe instruments designed to detect intentional exaggeration of symptoms or concerns when compared to established diagnostic norms (Rogers, 2008). The term *effort tests* is commonly used to describe measures designed to detect when patients give suboptimal effort, or perform to a degree less than what would be observed in someone who was unimpaired (Loser, 2013). The terms "malingering," "suboptimal effort," and "poor effort" are each used to describe a pattern of behavior in which participants exaggerated symptoms or underperformed in order to gain some sort of external reward or to avoid a painful or unpleasant consequence.

Possible Reasons for Sub-Optimal Effort

Measures of effort are primarily administered in forensic settings (Bauer et al., 2007). This may be because exaggeration of symptoms is more abundant in forensic contexts than in other contexts (Bush et al., 2005). For example, a criminal defendant may have his or her sentence extended if it is determined that he or she did not put forth maximum effort on a court-ordered psychological evaluation (Kucharski, Ryan, Vogt, & Goodloe, 1998). If a defendant can demonstrate an elevated pathology, it may lend credibility to certain defenses that excuse or justify his or her crime. Conversely, defendants may also deny pathology if they perceive that presenting themselves more favorably may decrease their served time.

In personal injury litigation, pain and suffering may entitle claimants to damages. Because pain is subjective, litigants and their attorneys typically have a great incentive to exaggerate symptoms on psychological tests (Mendelson & Mendelson, 2004). The ability to demonstrate an enduring effect as the result of an accident or injury directly impacts the litigants' potential settlement or award.

Studies have indicated that as many as 30% of disability claimants have been determined to be giving suboptimal effort or malingering during psychological evaluation (Mittenberg, Patton, Canyock, & Condit, 2002). Individuals attempting to claim government benefits may desire to do so in order to have a solution to socioeconomic problems. A consistent and predictable income, particularly if the claimant has some form of health concern that is not recognized as a disability, may feel justified. Antisocial acts or behaviors, career dissatisfaction, work conflict, and end of career concerns may also contribute to the reasons why disability claimants may exaggerate their physical and psychological functioning. Some claimants may also attempt to change a medical diagnosis to better fit the federal guidelines of a disability

Federal legislation requires colleges and universities to provide accommodations to students with diagnosed disabilities that interfere with the ability to function as a student (Gordon & Keiser, 1998). This is based, in part, on the principle of *equilibrium.* Referring to the American with Disabilities Act (ADA), equilibrium is the idea that students with diagnosed disabilities may need reasonable accommodations in order to perform at their true ability level (Gordon & Keiser, 1998). Such accommodations may include leniency with absences, private testing environments, access to note takers, and extended time for assignments (Lewandowski, 2014). Research has indicated that there is growing concern that college students may feign symptoms of disabilities, particularly of Attention-Deficit/Hyperactivity Disorder (ADHD), in order to gain access to such accommodations (Jasinkski et al., 2011). In a study by Larrabee (2012), students with documented disabilities noted that having a separate room, a scribe, a reader, and word processor were of more benefit than did students without documented disabilities. Additionally, a significant number of students without disabilities stated that they believed that all students should have access to accommodations, or that tests should be redesigned so that accommodations are not needed by any student. This demonstrates that a great number of students in colleges and universities view academic accommodations as beneficial.

Sub-Optimal Effort in College Students

There are multiple theories that espouse that low effort on testing procedures produces deleterious effects on test scores (Wise & DeMars, 2005). Underlying these theories is the idea that test scores are impacted by two factors: (a) expectancy, or the student's belief that he or she can perform the task, and (b) value, the belief that the student has regarding why he or she should complete the task. Additionally, Wise and DeMars add that students give a range of effort depending on their interpretation of the intrinsic value, utility, and perceived costs of completing

the testing. When examining college students' effort levels in relation to the presentation of their academic ability, it may be the case that students who give poor effort do so because they place high intrinsic value on accommodations, that it will get them closer to a goal, and that the benefit of giving poor effort outweighs the potential consequences.

Current studies that examine college student effort levels primarily use a diagnosis of ADHD as the focus of the studies (Musso & Gouvier, 2014). There has been an increase in the number of students that report ADHD symptoms in college settings (Schwarz, 2012). The diagnosis of ADHD relies heavily on self-report measures. Some research indicates that students seeking accommodations for ADHD may want prescribed medication (Sollman, Ranseen, & Berry, 2010). Studies report that as many as one-third of college students who were prescribed medication such as Adderall or Ritalin may divert that medication for one or more of the following purposes: recreational use, a source of income (resale), studying longer hours, or increasing concentration or ability to hyper-focus (Sollman et al., 2010). Additionally, with the number of job opportunities waxing and waning for college graduates, it may be the case that college students feel a great amount of pressure to be very successful. Students with other diagnoses such as depression or anxiety may feel the same pressures (Tan, Slick, Strauss, & Hultsch, 2002). As a result, they may perceive that academic accommodations will make them more competitive in the job market by increasing their grade point average (Suhr & Wei, 2013). These accommodations may include extended time on tests and assignments, assistance to respond, alternate testing locations, math aids, visual aids, direction clarification, and course waivers/substitution (Lai & Berkeley, 2012). In a study by The College Board, receiving additional time on tests was shown to increase Scholastic Aptitude Test (SAT I) scores in some cases by more than 100 points (Camara, Copeland, & Rothschild, 1998). Although a more recent

study indicated that additional time on tests should be used with caution (Mandinach, Bridgeman, Cahalan-Laitusus, & Trapani, 2005), low, medium, and high ability students are still seeking this accommodation (Katz, 2015). Additionally, it is extremely difficult to be admitted to college, graduate school, or to get a professional license without sitting for some sort of standardized or high-stakes test (GRE, LSAT, MCAT, etc.).These tests are directly linked to a person's ability to achieve educational and occupational goals, and are a large part of a student's calculations of future success. Hence, the incentive to perform well on these tests is high.

Detecting Sub-Optimal Effort

As a result of the potential benefits one could receive should they approach testing with sub-optimal effort, a number of measures have been developed that are designed to assess whether or not a client is giving full effort on psychological tests (Rosenfeld, Sands, & Van Gorp, 2000). These measures may take one of two forms: (a) they may be stand-alone measures, (i.e., overt tests specifically designed to detect effort) or (b) they may be embedded measures of effort (Babikian, Boone, Lu, & Arnold, 2006). Embedded measures are subtly integrated into a given test that has been designed for another purpose, such as assessing emotional difficulties. For example, the Minnesota Multiphasic Personality Inventory Second Edition (MMPI-2) is designed to assess a client's current level of psychological functioning across a number of dimensions such as depression or anxiety. However, the test also measures consistency and exaggeration in each response style. Thus, while the test has been created to assess emotional difficulties, the form of questions on the test also allows clinicians to notice when a client is attempting to appear in an unrealistically positive or negative light.

The decision to use stand-alone or embedded tests of effort may incorporate several dimensions (Schutte, Millis, Axelrod, & VanDyke, 2011). While a single stand-alone test may

generalize to an entire battery, it may be important to include effort measures throughout testing procedures in order to detect sub-optimal effort (Van De Kreeke, 2013). In a study of expert neuropsychologists who specialized in financial compensation and personal injury litigation claims, clinicians changed the testing protocol to include additional measures of effort when an examinee was thought to be giving suboptimal effort (Tan, Slick, Strauss, & Hultsch, 2004). As a result, clinicians responded in a variety of ways after incorporating additional measures of effort in the evaluation process. Most neuropsychologists encouraged clients to give better effort, administered additional effort tests, while other neuropsychologists directly confronted or warned the client, terminated assessment earlier than expected, contacted the referring attorney immediately, and included statements of invalidity in written reports.

Stand-alone effort tests typically produce higher psychometric face validity and accuracy, but may be susceptible to coaching influence (Victor, Boone, Serpa, Buehler, & Ziegler, 2009). Because stand-alone tests are commonly used in a number of test batteries, it is easier for non-clinicians to research, study, and do well on them. If a client can be coached to do well on a stand-alone effort test, but to do poorly on the psychological measures themselves, coaching may render the use of stand-alone measures useless. Alternatively, because clients may not perceive when embedded measures are assessing for consistency and exaggerated responses, they are often less susceptible to coaching. Embedded measures, however, may not be as accurate in distinguishing suboptimal effort from true psychopathology or cognitive impairment. Meyers, Volbrecht, Axelrod, and Reinsch-Boothby (2011) explain that when used appropriately with the populations they were designed for, embedded tests may be used both as a cognitive measure and a screen for effort. Consequently, although this convenience may add to the desirability of embedded measures, it may mean that clinicians will have to closely examine the

results of multiple measures to have confidence that the results reflect a relationship to effort and not cognitive impairment.

Much of the research that incorporates measures of effort focuses on primarily one form of effort test or the other. Some research has, however, examined the benefit of using either stand-alone or embedded measures in certain environments (Miele, Gunner, Lynch, & McCaffery, 2012). For example, Van De Kreeke (2013) looked at both embedded and stand-alone measures of effort in a population of criminals that had a formal diagnosis of malingering. Her study found that the stand-alone measures added a significant benefit to the test battery administered. In other words, the stand-alone measures added sensitivity and were able to detect malingerers that the embedded measures were not able to detect. In a study done with a military sample by Armistead-Jehle and Hansen (2011), whether or not a stand-alone measure was more effective than an embedded measure appeared to depend on environmental factors, such as rank or whether or not the participants were active duty. This may suggest that there are important confounds to consider when determining a patient's level of effort.

Statement of the Problem

Tests of effort are rarely used in non-forensic settings (Bauer et al., 2007). A small body of research exists that examines the use of embedded measures of effort and the use of stand-alone measures of effort. The research that does exist typically looks at forensic and military populations. Additionally, no meaningful and consistent body of research addresses the relationship between results of each type of effort test and academic ability. This information may assist colleges and universities in a number of ways. Because colleges and universities are required by federal legislation to provide accommodations for students with disabilities of various natures, and because college populations can be fairly transient, universities have a

strong interest in obtaining the most accurate information in the most efficient way possible (Victor, et al., 2009). Due to the rise in the amount of students that may feign disability in order to gain accommodations, colleges and universities would likely benefit by developing assessment protocols that would allow them to distinguish between students that genuinely need accommodative resources and students attempting to access resources undeservingly for personal gain.

Statement of Purpose

The purpose of this study is to add to the body of research that seeks to improve psychological assessment procedures. This study seeks to examine the relationship between tests of effort and academic ability. More specifically, this study seeks to examine the relationship between each type of test of effort, embedded and stand-alone, and measures of academic ability. It is hypothesized that indications of suboptimal effort will predict significantly lower scores on academic and cognitive measures. To wit, there is no significant difference between the correlations of each type of effort test and resulting academic scores.

Research Questions

This study addresses the following research questions:

1. What is the correlation between embedded effort tests and scores of academic ability in college students?
2. What is the correlation between stand-alone effort tests and scores of academic ability in college students?
3. Is there a significant difference between the correlation of embedded tests of effort and academic ability and stand-alone measures of effort and academic ability?

Method

Procedure

This study incorporated the use of a pre-existing data set. Data were collected at the University Accessibility Center (UAC) at Brigham Young University in Provo, Utah. The data set includes scores on a number of assessments for 602 students. Based on presenting concerns, not all students were administered all assessments. The data were aggregated, de-identified, and provided to the researcher for purposes of this study. Data were collected between the years of 2007 and 2014. Assessment protocols were stored in locked file cabinets and only accessible with a key that stayed locked in a drawer. Once data were compiled electronically, it was stored on a secure network drive with limited access. Only individuals granted permission could access the drive, after entering a username and password. The Institutional Review Board at Brigham Young University has indicated that due to the archival nature of the data, this study involves no contact with human subjects, and thus did not need to be monitored by the IRB at the university.

Participants

The data comprises test scores from students who presented at the UAC. Presenting concerns of participants varied, but each student was assessed for potential diagnoses that warrant accommodations in their University classes. The age of participants ranged from 16 to 61 years old, with the mean age being 24.9 years. Males comprised 38% of the participants, while females comprised 62% of the participants. No information regarding racial or ethnic background of the participants was available. All participants provided written consent that they were both aware that their test scores may be used for research purposes and that they were willing to allow their information to be used. No identifying information for any participant was included in the data set.

Measures of Effort

Data from the following measures were used to address the research questions:

Test of Memory Malingering (TOMM). The Test of Memory Malingering is a neuropsychological assessment designed to distinguish between clients with true memory impairments and clients that are feigning memory impairments (Tombaugh, 1996). The test is designed to screen for malingering via memory without also picking up other neurological impairments. The TOMM has two learning trials and an optional trial that assesses retention. The test uses visual stimuli, and signals potential malingerers by categorizing them as either below chance or by using criteria specific to clients with head injuries and cognitive impairments. The TOMM is considered the most widely used assessment of effort and malingering, and thus, has been the subject of numerous validation studies in a variety of contexts with different populations. One such extensive study concluded that performance on the TOMM by patients with traumatic brain injuries (both litigating and non-litigating) was comparable to that of cognitively intact individuals (Rees, Tombaugh, Gansler, & Moczynski, 1998). Therefore, individuals who do not perform well on the TOMM are suspect of exerting sub-optimal effort (Teichner & Wagner, 2004).

O'Bryant, Engel, Kleiner, Vasterling, and Black (2007) assert that the use of TOMM Trial 1 as a brief screening measure for insufficient effort yields adequate diagnostic accuracy, particularly when time is of the essence and a determination of insufficient effort is important yet not critical to the clinical question at hand. In that study, participants were administered only Trial 1 of the TOMM and then subsequently administered the full assessment. Individuals found to be giving poor effort were identified on both Trial 1 of the TOMM and the full administration. Diagnostic accuracy was determined by comparing cut scores from this study to expected values.

Researchers showed that previous research boasted the ability to correctly identify 93% of individuals giving poor effort, while this study correctly identified 91% of those giving poor effort when using a 10% base rate. Bauer et al. (2007) conducted a similar study with mild head-injury participants and found "impressive diagnostic accuracy" when using the TOMM Trial 1 as a screening measure for effort. A third study concluded that patients scoring 45 or greater on Trial 1 of the TOMM are not likely to be suspected of inadequate effort on their overall TOMM performance (Gavett, O'Bryant, Fisher, & McCaffrey, 2005). Because the current study considers the TOMM in conjunction with multiple other measures of effort, research supports the use and consideration of only Trial 1 of the TOMM as a stand-alone measure of effort in determining its potential relationship to academic ability variables.

Word Memory Test (WMT). The Word Memory Test asks clients to memorize a list of twenty word pairs. Examples of pairs may include 'pencil-pen' or 'pig-bacon.' It is a stand-alone measure specifically designed to measure a person's effort on psychological testing. Clients are shown the twenty word pairs at the rate of one pair every six seconds. This procedure is repeated and the client is shown the list for a second time. Clients then are administered the Immediate Recognition (IR) trial. Results place clients in the categories of "Pass," "Caution," or "Fail." The Word Memory Test has had extensive validation in clinical forensic settings. In a comprehensive study, researchers indicated that removal of data screened by the WMT and signaled as demonstrating poor effort, significantly affected the results of the study (Green, Lees-Haley, & Allen, 2002). The WMT has demonstrated the ability to discriminate effectively and efficiently between patients and claimants exhibiting poor effort (Hartman, 2002). Additionally, research supports the use of the Immediate Recognition section of the WMT as a brief screening tool for sub-optimal effort (Bauer et al., 2007). Although some research indicates that the WMT may be

highly specific but not sufficiently sensitive (Gervais, Rohling, Green, & Ford, 2004), other research indicates that the WMT boasts a sensitivity rate as high as 98.4%, indicating an extremely low probability of false positives (Green, Montijo, & Braukhaus, 2011).

Validity Indicator Profile (VIP). The VIP consists of 178 questions, 100 of which are non-verbal and 78 of which are verbal questions. The non-verbal (VIP Nonverbal) questions consist of picture puzzles in which clients are to choose the puzzle piece that completes the picture. The verbal questions (VIP Verbal) ask clients to choose the word most similar to the stem word given. Clients may not be told the name of the test, which may help decrease the likelihood of extreme symptom exaggeration. The Validity Indicator Profile examines the motivation and effort components of a client's test taking approach. A compliant response style means the client exhibited high motivation and high effort, and thus gave valid results. A careless response style means that the client demonstrated some motivation but poor effort. An irrelevant response style occurs when both motivation and effort are poor. A malingering response style indicates that the client had a high amount of motivation to appear impaired (Allington, 2014).

The Validity Indicator Profile has been used to determine levels of effort for a wide variety of populations. Cockshell and Mathias (2014) determined that the VIP was useful because it tested domains not affected by other disorders, in particular Chronic Fatigue Syndrome. In another study, researchers found that in cognitively impaired individuals with schizophrenia and schizoaffective disorders, the VIP may produce an increased amount of invalid profiles when compared to the TOMM, but that performance by these impaired individuals was consistent over environment, demonstrating that the VIP can still be appropriately used by this population. Drwal (2005) added that the VIP could be used to screen

for effort in a 15 to 18 year old population to a similar degree as an adult population. When used in conjunction with the TOMM, the VIP can increase incremental validity (Bayliss, 2014). Additionally, the VIP can adequately detect feigned ADHD and Reading Disorder (Frazier, Frazier, Busch, Kerwood, & Demaree, 2008). Because research does not support using cut-off scores for the VIP Verbal subtest, data for this section of the VIP was omitted from the current study.

California Verbal Learning Test 2nd Edition (CVLT-II). The CVLT-II measures both recall and recognition over a number of trials. The trials encompass both immediate recall components and delayed recall components. Clients hear a list of words and then recall words that they heard. This occurs five times. They are then given a different list of words and asked to immediately recall them. The test then consists of short-delay free recall trial and a short-delay cued recall trial of the original list. Clients then complete non-verbal testing during a 20-minute delay (Delis, Kramer, Kaplan, & Ober, 2000). After another series of long-delay recalls and a yes/no recognition trial, clients are then administered the Forced Choice Recognition Trial.

The Forced Choice Recognition Trial of the CVLT-II may be used to detect suboptimal effort in individuals reporting memory and learning concerns, and demonstrates "strong predictive value in positive findings of inadequate effort" (Root et al., 2006, p. 695). In a sample of individuals with traumatic head injury (THI), researchers found that the CVLT-II has a low false positive rate of 7.46%, indicating that it correctly identified most of the participants exhibiting sub-optimal effort, particularly when used in conjunction with other symptom validity tests (Baker, Donders, & Thompson, 2000). Research has also demonstrated that those determined to be giving sub-optimal effort had a generally lower mean on the CVLT-II than those who were determined to be giving adequate effort (Bauer, Yantz, Ryan, Warden, &

McCaffery, 2005). In that study, researchers applied WMT cut scores to the CVLT to distinguish between post–Traumatic Brain Injury (TBI) military participants giving optimal and suboptimal effort. Furthermore, the CVLT-II has been shown to be able to distinguish litigating individuals with mild head injury and non-litigating individuals with moderate and severe brain injury (Millis, Putnam, Adams, & Ricker, 1995). The aforementioned study demonstrates that the CVLT-II has sensitivity to motivation and effort because even individuals with severe brain injury can produce passing scores.

Reliable Digit Span. The WAIS-IV is an assessment tool that aims to measure the intellectual capacity and ability of clients between the ages of 16 and 90 (Wechsler, 2008). Individual subtests comprise indices helpful to clinicians. The indices on the WAIS-IV include a Verbal Comprehension Index, a Perceptual Reasoning Index, a Working Memory Index, a Processing Speed Index, and a Full Scale IQ score. Individuals' scores are standardized by comparing raw scores to established norms from similarly aged peers (Wechsler, 2008).

Reliable Digit Span examines and interprets raw data from the Digit Span subtest. It has been used repeatedly to assess client effort. Research indicates that the RDS is an appropriate embedded measure to screen for malingering (poor effort), boasting diagnostic accuracy upwards of 77% (Schroeder & Marshall, 2011). In a study with 141 college students, researchers found that the RDS correctly identified the vast majority of students not exhibiting poor effort. The same study found that this subtest rarely produced false positive results, meaning that less than four percent of subjects were incorrectly identified as displaying poor effort (Harrison, Rosenblum, & Currie, 2010). In a meta-analytic review of studies using RDS as a measure of screening for suboptimal effort, researchers found that the RDS effectively discriminated

between honest responders and dissimulators, with a moderate to high effect size (Jasinski, Berry, Shandera, & Clark, 2011).

Integrated Visual and Auditory Continuous Performance Test Advanced Edition. The IVA-AE is a computerized test that measures response control and attention. It consists of three stages – the Warm-Up Period, the Practice Period, and the Main Test. Clients are asked to click the mouse only when they see a "3" or hear a "5." Clients are instructed to NOT click when they see a "5" and hear a "3." Thus, the test requires clients to exhibit sustained and focused attention. The test may also provide clinical information that can be used to better understand a client's concerns related to attention that result from other medical problems, such as head injuries and dementia. Research has demonstrated that while ADHD self-report measures can be faked, the IVA-AE could not be faked on 81% of its scales and successfully distinguished malingerers from those with true impairment (Quinn, 2003). The Full Scale Attention Quotient (FSAQ) measures a person's ability to attend to a task by recording how many times an individual omits a response. The Full Scale Response Control Quotient (FSRCQ) measures an individual's impulsivity by assessing the number of times the person has an error of commission, or responds to a stimulus incorrectly (Alfano & Boone, 2007). Quinn's 2003 study demonstrated that analyzing the FSAQ and FSRCQ yielded sensitivity of 81% and specificity of 91%, indicating that there is a low probability of the measure committing either a Type I or Type II error. Additionally, the IVA-AE has been shown to have excellent sensitivity (92%) and specificity (90%), while also having adequate concurrent validity with other continuous performance tests for ADHD (Forbes, 1998).

Measures of Academic Ability

Ability is the natural aptitude or acquired proficiency necessary to complete a task (Merriam Webster, 2016). Academic ability refers to the capacity of a student to perform in areas of reading, writing, and mathematics. Academic ability can be assessed by a number of tools, including performance on aptitude tests and tests of achievement. In college students, measures of academic ability are often used to determine a student's potential for success in college. While cognitive measures assess a student's ability to learn concepts, tests of ability and achievement assess a student's knowledge in a given domain. Students' motivation to succeed in college may also play an important role in their academic ability. In a study conducted with university students in South Africa, researchers found that intrinsic motivation and effort were the strongest predictors of academic performance, as measure by Grade Point Average (Goodman et al., 2011). Measures of academic ability were selected based on established relationship to detection of effort.

Woodcock Johnson Test of Achievement 3rd Edition (WJ-III). The Woodcock Johnson Test of Achievement tests reading, mathematics, written language, oral language, and academic knowledge. There are 22 subtests that assess these five domains. Different combinations of the subtests provide helpful interpretable clusters that help clinicians form a more comprehensive picture of an individual's academic ability (Woodcock, McGrew, & Mather, 2001). The WJ-III has been frequently used as a measure to diagnose learning disabilities and academic achievement (Krasa, 2007).

Letter-word identification. Letter-Word Identification measures the ability to identify words. Students are asked to read letters and words aloud from a list, without context (Mather, Wendling, & Woodcock, 2001). A low score on this subtest may suggest lack of reading

vocabulary (Krull, et al., 2008). A lack of reading vocabulary may impact a college student in multiple ways, including understanding written instructions for assignments and tests. Remedies for these concerns typically include the use of assistive technology (i.e., computers and recorders during class), and extended time on assignments and tests.

Fluency scores. Academic Fluency is comprised of Reading Fluency, Math Fluency, and Writing Fluency, each of which is a timed subtest. Because the items increase in difficulty, they assess an individual's automaticity, or ability to respond to questions automatically with information that has been rehearsed and learned (Schrank & Flanagan, 2003). Automaticity in different academic arenas may be impaired in individuals with ADHD (Fabio, Castriciano, & Rondanini, 2015); therefore, academic fluency scores were included in the current study because of their relationship to the accurate assessment of ADHD. Reading Fluency assesses a student's ability to read simple sentences quickly. Students are given three minutes to read sentences and indicate whether they are true or false. Math Fluency assesses a student's ability to perform simple addition, subtraction, and multiplication operations quickly. For Math Fluency, students are given a response book and asked to solve as many of the problems as possible. Writing Fluency measures an individual's ability to formulate and write simple sentences quickly. For Writing Fluency, individuals are given three-word prompts for each sentence/item and are asked to write as many sentences as possible in seven minutes (Mather et al., 2001). In order to reduce experiment wise error, and because Letter-Word Identification provides sufficient information regarding an individual's reading ability, Reading Fluency scores were omitted from the data analysis in the current study. Additionally, there has been research to suggest that the analysis of specific learning disability should include Letter-Word Identification, Academic Fluency, Math Fluency, and Writing Fluency (Fredstrom, n.d.).

Wechsler Adult Intelligence Scale – Matrix Reasoning. Although it is a measure of cognitive ability, Matrix Reasoning has been shown to significantly correlate with verbal fluency task performance (Dugbartey et al., 1999). Matrix Reasoning is an untimed subtest of the Perceptual Reasoning Index. Perceptual Reasoning is intended to measure an individual's capacity to solve problems, organize thoughts, and examine rules and logical relationships. Matrix Reasoning asks clients to identify the missing picture in a matrix from five different options. The items on Matrix Reasoning assess visuospatial ability as well as simultaneous processing (Sobel, 2014). This subtest is especially resistant to cognitive impairment such as TBI. In cases where researchers examined the performance of patients with mild, moderate, and severe TBI, matrix reasoning performance was shown to be consistent across conditions (Carlozzi, Kirsch, Kisala, & Tulsky, 2015). As such, Matrix Reasoning was included in this study as a measure of nonverbal intelligence.

Data Analysis

Data were analyzed using a point-biserial correlation coefficient (r_{pb}). Point-biserial correlations are useful when determining relationships in which one variable is continuous and the other is dichotomous (O'Grady, 1977). In the current study, we sought to understand the relationship between tests of effort and academic ability. In this study, one of the research questions essentially asks whether or not individuals have given poor effort. We sought to answer this question as either "yes" or "no." Therefore, in the analysis of the data, it became necessary to dichotomize the data for the measures of effort. Because data for the measures of effort was continuous, cut scores were used to divide the data into a "HIT" and "NO HIT" dichotomy. HIT indicates that an individual score was at or below the value (established by the literature for each measure; see below) determined to signify sub-optimal effort. The NO HIT

condition indicates that an individual score did not fall below the predetermined cutoff, and was at or above norms that indicate normal effort.

In some situations, it may be appropriate to use a canonical correlation to determine the relationship between two sets of variables. Although that would have provided information regarding significant relationships, in this study it would have limited the applicability and reach of the results. Because a canonical correlation analyzes relationships by grouping sets of variables, the results would have necessitated that anyone seeking to gain information about test taking behavior group their variables in the exact same manner. Thus, the point-biserial correlation allows researchers to scrutinize the relationship between each individual measure of effort and each individual measure of academic ability.

On the TOMM a criterion cut score of 44 and below indicates poor effort (Tombaugh, 1996). On the WMT-IR trial, a cut score of 82.5 and below indicates poor effort (Green, n.d.). On the VIP Non Verbal subtest, a cut score of 75 and below signifies probable poor effort (Frazier et al., 2008). A score of 15 or less on the CVLT-II Forced Recognition Trial indicates poor effort (Moore, 2004). The cut off scores for the FSAQ and FSRCQ of the IVA-AE are 75 and below and 40 and below, respectively (Quinn, 2003). RDS scores less than or equal to seven signify poor effort (Babikian et al., 2006).

Writing Fluency, Math Fluency, Academic Fluency, and Letter-Word Identification are subtests that have a standard score mean of 100 with a standard deviation of 15. Scores 69 and below are classified as “very low,” scores 70-79 are classified as “low,” scores 80-89 are “low average,” scores between 90 and 110 are classified as “average,” scores 11-120 are “high average,” scores 121-130 are classified as “superior,” and scores at or above 131 are “very superior.” Matrix Reasoning has a scaled score mean of 10 with a standard deviation of three.

Scores one through seven are classified as “below average,” scores eight through 12 are classified as “average,” while scores 13-19 are classified as “above average.”

Results

This study sought to examine whether or not there is a meaningful relationship between two different types of effort tests (stand-alone and embedded) and four different measures of academic ability (Letter-Word Identification, Academic Fluency, Writing Fluency, and Math Fluency) and one measure of cognitive ability (Matrix Reasoning). Because Matrix Reasoning has been highly correlated with verbal fluency task performance (Dugbartey et al., 1999), for the purposes of this study it has been grouped with the other measures of academic ability in the following analyses. Data were gathered at Brigham Young University’s Accessibility Center between 2007 and 2014. Testing protocols were individualized, resulting in different sample sizes for each measure. Descriptive statistics are included for the measures of effort and the measures of academic ability (see Table 1 and Table 2).

Table 1

Descriptive Statistics for Stand-Alone and Embedded Measures of Effort

Test	N	Hit Frequency (%)	Minimum Standard Score[a]	Maximum Standard Score[b]	Mean	Std. Deviation
TOMM Trial 1	232	12.1	26.0	50.0	48.034	3.4801
WMT IR	257	8.6	15.0	100.0	95.243	9.3444
VIP Total Nonverbal	32	12.5	59.0	97.0	84.875	9.4621
Reliable Digit Span	29	21.6	0	16.0	9.021	2.0651
CVLT-II Forced Choice	18	16.7	8.0	16.0	15.056	2.5776
FSRCQ	39	28.2	26.0	119.0	86.667	23.6280
FSAQ	39	10.3	0	115.0	81.359	31.2026

Note. Sample sizes for tests varied based on individualized testing procedures.
TOMM = Test of Memory Malingering; WMT IR = Word Memory Test Immediate Recall; VIP = Validity Indicator Profile; CVLT = California Verbal Learning Test; FRSCQ = Full Scale Response Control Quotient; FSAQ = Full Scale Attention Quotient.
[a] Represents the lowest value in data set for test
[b] Represents highest value in data set for test

Table 2

Descriptive Statistics for Measures of Academic Ability

Ability Test	N	Minimum Standard Score[a]	Maximum Standard Score[b]	Mean	Std. Deviation
Academic Fluency	468	13.0	146.0	94.528	13.6904
Letter-Word Identification	464	10.0	137.0	100.149	10.7479
Math Fluency	477	10.0	146.0	89.017	13.8676
Matrix Reasoning	85	6.00	18.00	13.6706	2.41714
Writing Fluency	463	3.0	149.0	99.0	13.2420

Note. Sample sizes for tests varied based on individualized testing procedures.
[a] Represents the lowest value in data set for test
[b] Represents highest value in data set for test

Stand-Alone Measures

A point biserial correlation was conducted using the dichotomized stand-alone variables (TOMM, WMT IR, VIP Nonverbal) and the measures of academic ability (Writing Fluency, Matrix Reasoning, Math Fluency, Letter Word Identification, and Academic Fluency) to evaluate the strength of the relationship between the two variable sets (see Table 3). Each stand-alone variable was dichotomized in order to indicate whether or not an individual score was above or below published cutoffs relating to sub-optimal effort. VIP Nonverbal was found to have a significant relationship with Writing Fluency (r_{pb} = -.364, p<.05). TOMM was found to have a significant relationship with Writing Fluency (r_{pb} = -.203, p<.01).

Table 3 *Correlations for Stand-Alone Measures and Academic Ability Measures*

Measure	Correlations	Academic Fluency	Letter-Word Identification	Math Fluency	Matrix Reasoning	Writing Fluency
TOMM_Hit	Pearson Correlation	-.027	.017	-.086	.157	-.203**
	Sig. (2-tailed)	.696	.808	.208	.302	.003
	N	210	204	215	45	209
WMTIR_hit	Pearson Correlation	.052	.085	.111	-.308	-.008
	Sig. (2-tailed)	.434	.194	.089	.285	.899
	N	232	232	236	14	234
VIPNONVERBAL_hit	Pearson Correlation	-.277	-.348	-.115	.c	-.364*
	Sig. (2-tailed)	.125	.051	.531	.	.041
	N	32	32	32	0	32

Note. Measure variables were dichotomized based on established cut scores. An individual's performance above the cut score was coded as 1, performance below the cut score was coded as 0. For all scales, lower scores are indicative of poor effort. TOMM = Test of Memory Malingering; WMT IR = Word Memory Test Immediate Recall; VIP = Validity Indicator Profile.

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

c. Cannot be computed because at least one of the variables is constant.

Embedded Measures

A point biserial correlation was run to determine the relationship between dichotomized embedded variables and measures of academic ability (see Table 4). RDS was shown to have a significant relationship with Academic Fluency (r_{pb} = -.235, p <.01), Letter-Word Identification (r_{pb} = -.261, p < .01), Math Fluency (r_{pb} = -.214, p < .01), and Writing Fluency (r_{pb} = -1.78, p <.01).

Table 4

Correlations Between Embedded Effort Measures and Academic Ability Measures

Test	Correlations	Academic Fluency	Letter-Word Identification	Math Fluency	Matrix Reasoning	Writing Fluency
FSAQ_hit	Pearson Correlation	-.153	.036	-.097	.033	-.059
	Sig. (2-tailed)	.360	.834	.573	.866	.733
	N	38	36	36	28	36
FSRCQ_hit	Pearson Correlation	.019	.115	.035	.164	-.032
	Sig. (2-tailed)	.910	.506	.840	.405	.855
	N	38	36	36	28	36
CVLT_hit	Pearson Correlation	-.199	-.284	-.104	.[a]	-.123
	Sig. (2-tailed)	.443	.269	.691	.	.639
	N	17	17	17	0	17
RDS_hit	Pearson Correlation	-.235**	-.261**	-.214**	-.214	-.178**
	Sig. (2-tailed)	.000	.000	.000	.053	.002
	N	302	295	305	82	295

Note. FSAQ = Full Scale Attention Quotient; FSRCQ = Full Scale Response Control Quotient; CVLT = California Verbal Learning Test; RDS = Reliable Digit Span.
**. Correlation is significant at the 0.01 level (2-tailed).
a. Cannot be computed because at least one of the variables is constant.

Discussion

The purpose of this study was to add to the body of research that seeks to improve neuropsychological assessment procedures. This study sought to examine the relationship between tests of effort and academic ability. More specifically, this study sought to examine the relationship between both types of test of effort, embedded and stand-alone, and measures of academic ability. Examining the relationship between types of tests of effort and academic ability would help test administrators determine if students are potentially feigning impairments

related to their expected knowledge in academic domains. We hypothesized that indications of suboptimal effort would predict significantly lower scores on academic measures. In other words, a HIT on a measure of effort would be significantly negatively correlated with measures of academic ability.

Sample sizes for each individual variable varied. Each participant in the data set was administered an individualized protocol, meaning not all tests were administered to each participant. Consequently, the sample size for each measure was different and created missing data points. We analyzed the data using point-biserial correlations despite the missing data points and varied sample sizes.

Two of the three stand-alone measures had significant correlations with Writing Fluency. Confidence in the results of this study rely on two primary factors: (a) the strength of the construct validity of the measures used, and (b) the reasons why the correlations were significant with Writing Fluency. In order for the results of this study to have meaning, it must be the case that both the TOMM and the VIP measure what they purport to measure. If neither measure actually assessed the construct of effort, the significant correlation with Writing Fluency may have been due to some undetermined construct. There may be some worry that in this study that we are possibly measuring a cognitive construct. For example, some research indicates that it is unlikely that credible individuals with true neurologic or psychiatric impairments can pass all embedded effort tests (Victor, Boone, Serpa, Bueler, & Ziegler, 2009). It is important to note, however, that even this research acknowledges the utility of effort tests and indicates that clinicians should only adjust their interpretation of results instead of dismissing them. Thus, it is necessary to understand the normative data for each of these stand-alone tests. In the development of both the TOMM and the VIP, coached malingerers and TBI patients were

administered the tests. In each case, individuals with mild to moderate brain damage still provided valid results of the TOMM and the VIP. Thus, the development of these tests leads us to believe that level of cognitive functioning is not confounded with effort. Furthermore, factor analysis has demonstrated that memory and effort are distinct and separate constructs measured by these tests (Heyanka et al., 2015).

It is of note that the TOMM and VIP Nonverbal correlated only with Writing Fluency and no other measure of academic ability. This is likely due, at least in part, to the fact that writing involves more processes than other academic tasks. Writing Fluency tasks require planning, monitoring, reviewing, retrieving, and transcribing (Abdel Latif, 2013). More involved academic tasks such as Writing Fluency may be more susceptible to students giving poor effort because it is more physically and mentally demanding. As such, clinicians may want to more critically examine students' Writing Fluency scores when students fail on stand-alone measures of effort.

In this study, RDS was the only embedded measure of effort to demonstrate a significant relationship with measures of academic ability, demonstrating a strong consistency by correlating with four of the five measures of academic ability (Letter-Word Identification, Writing Fluency, Math Fluency, and Academic Fluency). This is in line with past research. In a study of 207 Veterans, researchers concluded that scores on verbal measures cannot be assumed to reflect actual ability levels in the face of non-credible effort (Sawyer, Young, Roper, & Rach, 2014). In that study, participants were divided into credible and non-credible groups using RDS. Once administered the WAIS IV, WRAT-4, CVLT-II, and WMS-IV, analyses revealed that non-credible effort had a moderate to large effect size on reading test performance, even when controlled for educational level. Our study expands upon this finding because in this study RDS

predicts more than just reading test performance, it predicted Math Fluency, Writing Fluency, Academic Fluency, and Letter-Word Identification. Additionally, whereas the study performed by Sawyer et al. used a highly niche and specialized population, this study involved a broader population, increasing its potential genralizability.

This consistency has the potential to inform testing procedures in college counseling centers. Although measures of effort are not typically given in colleges, many programs may use tests in which RDS is already embedded. The ability to use RDS to critically examine academic ability measures may save time, money, and may reduce the strain on university resources. Furthermore, many universities require students to get testing for learning disabilities outside of the university itself and provide documentation to the disability services office in order to get accommodations. Traditionally, there is no evidence of effort testing provided in this documentation. The results of this study indicate that there is a cost and time effective alternative that may be more resourceful for both university personnel and students. Additionally, no previous research has demonstrated a significant link between RDS and multiple academic ability measures.

Limitations

In psychological testing procedures, any number of factors may influence a single testing result (Sapp, 2002). Changes in method of administration, environment, client feeling and mood, and level of test anxiety can influence test results (Domino & Domino, 2006). Additionally, a small sample size means that any individual score has more influence on the result, which is why sample sizes as large as possible are ideal (Haebara, 1986). In the current study, tests were administered across seven years by a number of individuals with varying levels of experience. Students tested varied in age, gender, and diagnosis. Although all testing was supervised and

likely administered consistent with developer guidelines, any anomalies would not be known in this study because it is an archival de-identified data set.

The use of cut-scores presented challenges in the current study. Although there was research supporting the cut-scores used, on a theoretical note, there may not be a significant difference between an individual scoring above or below the cut score by a single point. In this study, adjusting the cut-scores, even minutely, may have changed the overall significance patterns and whether or not any measure of effort correlated with any measure of academic ability. There may be cause to assert that a HIT/NO HIT dichotomy is too rigid, and future research should examine whether or not a third category (not fail, but also not passed) is warranted. Some research suggests that there exists a population of individuals who have adequate cognitive functioning, no history of TBI, and still fail an effort test, and that cut-scores may not be appropriate in assessing these individuals (Willis, Farrer, & Bigler, 2011). Examiners should deeply consider their use of cut-scores and the environment in which they work. Some environments, such as universities, may have the luxury of fluidity when it comes to determining whether or not a student is giving poor effort. Forensic environments, however, more often rely upon stringent determinations to make critical decisions, and therefore may use cut-scores to communicate more definitive messages about clients. Although not likely to exist, a known-groups comparison in a university setting may help determine the true nature of the relationship between measures of effort and measures of academic ability. This would allow researchers to more definitively connect the constructs of effort and academic ability.

A significant limitation of this study is that the nature of the relationship between RDS and the measures of academic ability cannot be determined. The same can be said for the relationship between the VIP and Letter-Word Identification and Writing Fluency, although the

relationship between those variables is somewhat more intuitive because they are all verbal/vocabulary-based tests. It is not known whether or not low scores on the measures of effort is actually representative of malingering. While the VIP considers motivation and effort in its indices, there was not a significant relationship between either section of the VIP and most of the measures of academic ability. At best, the results of this study advocate for consideration of the VIP, RDS, and TOMM in testing procedures at university accessibility centers among accommodation seeking populations.

Furthermore, each measure of effort in its development defines ideal and suboptimal effort in slightly different ways. Consequently, the tests may get at slightly different constructs. If this is the case, the true nature of the relationship between measures of effort and any other variable may be difficult to determine.

It was expected that Matrix Reasoning would correlate with stand-alone and embedded measures of effort. In the current study, however, Matrix Reasoning did not correlate with a single test of effort. This may be because although matrix reasoning is resistant to influences of TBI and ADHD, it may not be sufficiently related to academic ability. Matrix reasoning provides information about cognitive process, which is a different construct than academic ability. Thus, it may be the case that performance on other measures of effort which require a greater cognitive load may predict matrix reasoning. Future research should examine the relationship between cognitive processing measures and academic ability measures.

Methodological limitations. Although the VIP was the test administered the least amount of times, the researcher does not have any indication as to why this may be the case. Because measures of effort are rarely given in non-forensic settings, those in charge of choosing assessment protocols may not be aware of the need or benefit of effort testing. Additionally,

because other stand-alone measures were administered more frequently than the VIP, test administrators may see little utility in adding data from the VIP in the absence of indication of poor effort from the TOMM or WMT. These tests are administered quite frequently and have built strong reputations in the psychological community, whereas little research examines the role of using the VIP in university settings.

Academic ability was operationalized to constitute a synthetic variable made of several criterion variables. Although research supports the use of these variables in determining effort, there are a large number of measures that exist to give information regarding academic ability, including Grade Point Average. GPA was not available to this researcher, and as such, future research should examine whether the predictor variables in this case have significant relationships with other variables measuring academic ability.

Archival data were presented to the researcher de-identified. The data were collected summarily for documentation purposes. It is possible that conducting a specific study regarding measures of effort and academic ability variables will lead to increased sample sizes of the variables, thus yielding different results.

Population biases. All data were collected from accommodation seeking students at BYU in Provo, Utah. While most schools have policies against academic dishonesty, Brigham Young University has a heavily enforced Honor Code, which promotes compliance with set rules. Additionally, the vast majority of students at the university identify as members of the Church of Jesus Christ of Latter-day Saints (typically over 95%; enrollment data obtained April, 2016), a religious sect that doctrinally supports honesty in all dealings with peers.

Implications

In environments where measures of academic ability are given, and there are questions of effort, it may be the case that RDS, TOMM, and VIP Nonverbal will give examiners a better understanding of clients' test taking behavior when used in conjunction with the results of other effort tests. The RDS is part of widely administered assessment tools; the WAIS-IV. As a result, it would require little extra time or efforts for administrators to examine whether or not RDS scores are consistent with the overall clients' data. Additionally, it may be the case that those in college settings may wish to focus upon the VIP Nonverbal as a measure of effort, despite the TOMM and WMT being widely used and accepted. If there are conflicting test results regarding effort, RDS results may serve to help examiners understand whether or not the student is accurately portraying his or her academic abilities.

A finding of no significance is informative despite limited generalizability. In the current case, although not every effort test was significantly correlated with academic ability measures, it still provides helpful information. For example, test administrators may want to consider administering the VIP as their primary assessment tool of effort instead of continuing to use it as a tertiary screening of effort. Despite the TOMM and WMT being more widely used instruments in the industry, this research indicates that the VIP may be a more helpful instrument with this population when examining the relationship between effort and academic ability.

Additionally, it may be the case that practitioners may need to develop a sophisticated, yet covert testing protocol procedure. It may not be efficacious to administer tests of effort solely in the beginning of testing. Because each protocol is individualized, practitioners may wish to conduct clinical interviews and administer personality testing prior to deciding when effort tests should be administered for each student.

Summary and Conclusion

Tests of effort are measures that seek to determine whether or not a client is answering questions in a manner consistent with the developer's guidelines. Sub-optimal effort occurs when a client does not perform up to his or her capacity with the intent of skewing test results. Individuals may do this for a number of reasons, including attempts to gain disability or military benefits, to appear impaired for litigation purposes, or to receive accommodations for university classes, or to perform well on high stakes testing such as the GRE, MCAT, etc., or professional licensing exam. The types of tests that assess for effort include embedded and stand-alone measures. Embedded measures are assessment tools that primarily assess for some other psychological construct, such as emotional stability, but covertly measure the client's effort and consistency. Stand-alone measures are marketed as tools that assess for malingering and poor effort.

This study sought to determine whether or not both types of effort measures had significant relationships with measures of academic ability. The research question served to contribute to the body of research that seeks to improve psychological testing procedures. It was hypothesized that a low score on a measure of effort would correlate with low academic ability scores. A point biserial correlation was run and determined significant relationships between RDS and Academic fluency, RDS and Math fluency, RDS and Writing Fluency, and RDS and Letter-Word Identification. Additionally, the VIP Nonverbal also correlated with Writing Fluency. The TOMM also correlated with Writing Fluency. These results indicate that university accommodation offices may wish to consider the RDS, TOMM, and VIP, particularly when there are questions of effort and test taking behavior.

Limitations of this study included a lack of information that could inform about the

nature of the relationship between variables. Additionally, most embedded variables did not appear to have a significant connection to academic ability variables. It is also important to note that the sample sizes for each measure varied due to individualized testing protocols. Lastly, the population used may be prone to being more honest than is typical of college students because of the university's honor code.

Future research should be conducted at a variety of universities in order to get a wider distribution of the types of students seeking accommodations. It would be useful to be able to have a population of known malingerers with whom to compare results, although this may not exist in university settings.

References

Abdel Latif, M. M. (2013). What do we mean by writing fluency and how can it be validly measured? Applied Linguistics, *34*, 99-105. doi:10.1093/applin/ams073

"Ability." 15 Apr. 2016. *Merriam-Webster.com*. Merriam-Webster, n.d. Web.

Alfano, K., & Boone, K. B. (2007). The use of effort tests in the context of actual versus feigned attention-deficit/hyperactivity disorder and learning disability. In K. B. Boone, K. B. Boone (Eds.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 366-383). New York, NY: Guilford Press.

Allington, J. P. (2014). *The psychometric properties of the validity indicator profile*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. (Order No. 3604665)

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

Armistead-Jehle, P., & Hansen, C. L. (2011). Comparison of the repeatable battery for the assessment of neuropsychological status effort index and stand-alone symptom validity tests in a military sample. *Archives of Clinical Neuropsychology, 26*, 592-601.

Babikian, T., Boone, K., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*, 145-159.

Baker, R., Donders, J., & Thompson, E. (2000). Assessment of incomplete effort with the California Verbal Learning Test. *Applied Neuropsychology, 7*, 111-114. doi:10.1207/S15324826AN0702_8

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the Test of Memory Malingering trial 1 and Word Memory Test immediate recognition as screening tools for insufficient effort. *Assessment, 14*, 215-222.

Bauer, L., Yantz, C. L., Ryan, L. M., Warden, D. L., & McCaffrey, R. J. (2005). An examination of the California Verbal Learning Test II to detect incomplete effort in a traumatic brain-injury sample. *Applied Neuropsychology, 12*, 202-207. doi:10.1207/s15324826an1204_3

Bayliss, C. M. (2012). *The detection of malingered intellectual developmental disorder*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. Retrieved from http://search.proquest.com/docview/1318907056?accountid=4488 (Order No. 3536723)

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., & National Academy of Neuropsychology Policy and Planning Committee. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419-426.

Carlozzi, N. E., Kirsch, N. L., Kisala, P. A., & Tulsky, D. S. (2015). An examination of the Wechsler Adult Intelligence Scales, Fourth Edition (WAIS-IV) in individuals with complicated mild, moderate and severe Traumatic Brain Injury (TBI). *The Clinical Neuropsychologist*, *29*, 21-37. doi:10.1080/13854046.2015.1005677

Camara, W. J., Copeland, T., & Rothschild, B. (1998). Effects of extended time on the SAT I: reasoning test score growth for students with learning disabilities. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-1998-7-effects-extended-time-sat-reasoning-students-learning-disabilities.pdf

Carroll, L. (2011, May 17). 'Steroids for school': College students get hooked on 'smart drugs'. Retrieved February 7, 2013, from http://www.today.com/id/43050779/site/todayshow/ns/today-today_health/t/steroids-school-college-students-get-hooked-smart-drugs/#.URRNoJjnGzA

Cockshell, S. J., & Mathias, J. L. (2014). Cognitive functioning in people with chronic fatigue syndrome: A comparison between subjective and objective measures. *Neuropsychology*, *28*, 394-405. doi:10.1037/neu0000025

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *CVLT-II: California Verbal Learning Test Manual*. The Psychological Corporation.

Detection of Malingering in Disability Evaluations. (2005, December 1). Retrieved September 9, 2015, from http://primarypsychiatry.com/determination-of-malingering-in-disability-evaluations/

Domino, G., & Domino, M. L. (2006). Psychological testing. Cambridge, GB: Cambridge University Press. Retrieved from http://www.ebrary.com

Drwal, W. R. (2005). *The use of the Validity Indicator Profile with a 15- to 18-year-old population.* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. (305360521). Retrieved from http://search.proquest.com/docview/305360521?accountid=4488 (Order No. 3164789).

Dugbartey, A. T., Sanchez, P. N., Rosenbaum, J. G., Mahurin, R. K., Davis, J. M., & Townes, B. D. (1999). WAIS—III Matrix Reasoning Test performance in a mixed clinical sample. *Clinical Neuropsychologist*, *13*, 396-404. doi:10.1076/1385-4046(199911)13:04;1-Y;FT396

Fabio, R. A., Castriciano, C., & Rondanini, A. (2015). ADHD: Auditory and visual stimuli in automatic and controlled processes. *Journal of Attention Disorders*, *19*, 771-778. doi:10.1177/1087054712459562

Forbes, G. B. (1998). Clinical utility of the Test of Variables of Attention (TOVA) in the diagnosis of attention-deficit/hyperactivity disorder. *Journal of Clinical Psychology*, *54*, 461-476. doi:10.1002/(SICI)1097-4679(199806)54:4<461::AID-JCLP8>3.0.CO;2-Q

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology*, *23*, 501-509. doi:10.1016/j.acn.2008.04.001

Fredstrom, J. (n.d.). Which WJ-III Subtests Should I Administer? Retrieved from http://www.asec.net/VideoTrainings/Woodcock-Johnson Achievement Test Cluster Scores needed for SLD.pdf

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*, 1-4. doi:10.1207/s15324826an1201_1

Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., & ... Schlechter, A. (2011). An investigation of the relationship between students' motivation and academic performance as mediated by effort. *South African Journal of Psychology*, *41*, 373-385. doi:10.1177/008124631104100311

Gordon, M. & Keiser, S. (1998). *Accommodations in higher education under the Americans with Disabilities Act (ADA).* New York, NY: Guilford Press.

Green, P. (n.d.) *The pervasive influence of effort on neuropsychological tests*. Unpublished manuscript. Retrieved from: http://wordmemorytest.com/WMT/Neuropsych_scores_effort.doc

Green, P., Lees-Haley, P. R., & Allen, L. I. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*, 97-124. doi:10.1300/J151v02n03_05

Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the word memory test and medical symptom validity test in groups with severe verbal memory impairment. *Applied Neuropsychology, 18*, 86-94. doi:10.1080/09084282.2010.523389

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

Haebara, T. (1986). Tables for determining sample size in correlational research. *Japanese Journal of Educational Psychology*, *34*, 155-158.

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, *17*, 283-293. doi:10.1177/1073191109348590

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the word memory test. *Archives of Clinical Neuropsychology, 17*, 709-714.

Heyanka, D. J., Thaler, N. S., Linck, J. F., Pastorek, N. J., Miller, B., Romesser, J., & Sim, A. H. (2015). A factor analytic approach to the validation of the Word Memory Test and Test of Memory Malingering as measures of effort and not memory. *Archives of Clinical Neuropsychology*, *30*, 369-376. doi:10.1093/arclin/acv025

High Tech Center Training Unit. (n.d.) Retrieved from: http://www.htctu.fhda.edu/publications/articles/ld/achieve.pdf.

Jasinski, L. J., Berry, D. T. R., Shander, A. I., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale digit span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology, 33*, 300-314.

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

Katz, S. (2015, July 27). The clock is ticking: Extended time and ADHD. [Web log post]. Retrieved from http://www.neuropl.us/blog/uncategorized/the-clock-is-ticking-extended-time-and-adhd/

Krasa, N. (2007). Is the Woodcock-Johnson III a test for all seasons? Ceiling and item gradient considerations in its use with older students. *Journal of Psychoeducational Assessment*, *25*, 3-16. doi:10.1177/0734282906291768

Krull, M., Norton, S., Reither, L., Flores-Charter, P., Hukill-DeRock, T., Tunstall, C… De La Cerda, K. (2008). *Selecting software for students with learning disabilities.* Retrieved from http://www.htctu.net/trainings/manuals/tutorials/ldsoft/document/sssld_d.pdf

Kucharski, L. T., Ryan, W., Vogt, J., & Goodloe, E. (1998). Clinical symptom presentation in suspected malingerers: An empirical investigation. *Journal of the American Academy of Psychiatry and the Law, 26*, 579-585.

Lai, S. A., & Berkeley, S. (2012). High-stakes test accommodations: Research and practice. *Learning Disability Quarterly*, *35*, 158-169.

Larrabee, G. J. (2012). Assessment of malingering. In G. J. Larrabee, G. J. Larrabee (Eds.), *Forensic neuropsychology: A scientific approach (2nd ed.)* (pp. 116-159). New York, NY: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625-631.

Lewandowski, L., Lambert, T. L., Lovett, B. J., Panahon, C. J., & Sytsma, M. R. (2014). College students' preferences for test accommodations. *Canadian Journal of School Psychology, 29*, 116-126.

Loser, N. M. (2012). *Malingering detection measure utility and concordance in a university accommodation-seeking student population.* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. Retrieved from http://search.proquest.com/docview/1151806746?accountid=4488 (Order No. 3542627).

Mandinach, E. B., Bridgeman, B., Cahalan-Laitusus, C., & Trapani, C. (2005). The impact of extended time on SAT performance. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2005-8-impact-extended-time-sat-test-performance.pdf

Mather, N., Wendling, B. J., & Woodcock, R. W. (2001). *Essentials of WJ III tests of achievement assessment*. Hoboken, NJ: John Wiley & Sons Inc.

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded Symptom Validity Tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology*, *26*, 8-15. doi:10.1093/arclin/acq083

Mendelson, G. & Mendelson, D. (2004). Malingering pain in the medicolegal context. *Clinical Journal of Pain, 20,* 423-432.

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology, 27*, 10-22.

Millis, S. R., Putnam, S. H., Adams, K. M., & Ricker, J. H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment, 7*, 463-471. doi:10.1037/1040-3590.7.4.463

Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 3*, 227-240.

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*, 186-201.

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*, 511-521. doi:10.1080/13854040600611368

O'Grady, K. E. (1977). A program to compute point-biserial coefficients of correlation. *Educational and Psychological Measurement*, *37*, 263-265. doi:10.1177/001316447703700135

Osmon, D. C., Plambeck, E., Klein, L., & Mano, Q. (2006). The word reading test of effort in adult learning disability: A simulation study. *The Clinical Neuropsychologist, 20*, 315-324.

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*, 379-395. doi:10.1016/S0887-6177(02)00150-6

Rees, L. M., Tombaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memory Malingering (TOMM). *Psychological Assessment*, *10*, 10-20. doi:10.1037/1040-3590.10.1.10

Religions Represented at BYU. (n.d.). Retrieved from http://yfacts.byu.edu/Article?id=97&_sm_au_=iVVjqPs8Pv4fnJnq

Rogers, R. (2008). Clinical assessment of malingering and deception. New York, NY: Guilford Press.

Root, J. C., Robbins, R. N., Chang, L., & Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test-second edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society, 12*, 688-696.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*, 349-359.

Sapp, M. (2002). *Psychological and educational test scores: What are they?* Springfield, IL: Charles C. Thomas Publisher.

Sawyer, R. I., Young, J. C., Roper, B. L., & Rach, A. (2014). Are verbal intelligence subtests and reading measures immune to non-credible effort? *The Clinical Neuropsychologist*, *28*, 756-770. doi:10.1080/13854046.2014.920922

Schrank, F. A. (2010, June 18). Woodcock Johnson III Test of Achievement. Retrieved April 14, 2016, from http://www.iapsych.com/articles/schrank2010ip.pdf

Schrank, F. A., & Flanagan, D. P. (2003). *WJ III clinical use and interpretation: Scientist-practitioner perspectives*. Amsterdam, Netherlands: Academic Press

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist, 25,* 437-453.

Schutte, C., Millis, S., Axelrod, B., & VanDyke, S. (2011). Derivation of a composite measure of embedded symptom validity indices. *The Clinical Neuropsychologist, 25*, 454-462

Schwarz, A. (2012, June 9). Risky rise of the good-grade pill. Retrieved from http://www.nytimes.com/2012/06/10/education/seeking-academic-edge-teenagers-abuse-stimulants.html?pagewanted=all&_r=1

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: a survey of experts' practices. *Archive of Clinical Neuropsychology, 19*, 465-473.

Sobel, N. K. (2014). *Detection of performance invalidity in neuropsychological assessments: Utility of the matrix reasoning subtest* Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. Retrieved from http://search.proquest.com/docview/1459765810?accountid=4488 (Order No. 3598818)

Sollman, M. J., Ranseen, J. D., & Berry, D. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*, 325-335. doi:10.1037/a0018857

Suhr, J., & Wei, C. (2013). Symptoms as an excuse: Attention deficit/hyperactivity disorder symptom reporting as an excuse for cognitive test performance in the context of evaluative threat. *Journal of Social and Clinical Psychology, 32*, 753-769.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*, 495-505. doi:10.1076/clin.16.4.495.13909

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*, 455-464.

Tombaugh, T. N. (1996). *TOMM. Test of Memory Malingering*. New York, NY: Multi-Health Systems.

U.S. Office of the Inspector General. (2013, September). *Congressional response report on the Social Security Administration's policy on symptom validity tests in determining disability claims.* (Publication No. A-08-13-23094). Retrieved from https://oig.ssa.gov/sites/default/files/audit/full/pdf/A-08-13-23094.pdf

Van De Kreeke, D. (2013). *The examination of effort tests: Identifying an efficient approach to the assessment of malingering*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. http://search.proquest.com/docview/1458231478?accountid=4488 (Order No. 3575092)

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23*, 297-313.

Wechsler, D. (2008). Wechsler adult intelligence scale: Fourth edition (WAIS-IV) [Assessment instrument] San Antonio, TX: Pearson.

Willis, P. F., Farrer, T. J., & Bigler, E. D. (2011). Are effort measures sensitive to cognitive impairment? *Military Medicine*, *176*, 1426-1431. doi:10.7205/MILMED-D-11-00168

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*, 1-17. doi:10.1207/s15326977ea1001_1

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III. Itasca, IL: Riverside Publishing.

APPENDIX A:

Review of the Literature

Psychological Test Selection

Psychological tests are used primarily for making decisions about people (Urbina, 2014). When deciding whether or not engaging in psychological testing is appropriate, clinicians must first decide (a) the information sought from testing, (b) how the information will be used, (c) what other tools and sources of information are available, and (d) the benefit of using psychological tests over other sources of information. Clinicians must also decide if the testing process will contribute efficiency and utility to the answering of the clinical question being presented. When selecting psychological tests to include in a protocol, psychologists must consider the reliability and validity of the instrument (Heilbrun, 1992). Additionally, it is important to consider the accessibility of the development of the instrument. Forensic testing results which may be used in court will need to consider the degree to which each testing instrument can stand up to the court's rules of evidence (e.g., Neff, 2015; Heilbrun, 1992; Reynolds & Horton, 2012).

Although incentives to perform well on tests include potentially being placed in gifted programs (Jung & Gross, 2014), detection of suboptimal effort is a critical element of any psychological assessment (e.g., Bush et al., 2005; Larrabee, 2012). Without tests of effort, many clinicians consider neuropsychological batteries to be incomplete (Bauer et al., 2007). In order to express confidence in the test scores, diagnoses, and treatment recommendations, there must be confidence that the tests were both administered and taken consistent with developer guidelines (Bush et al., 2005). Green, Rohling, Lees-Haley, and Allen (2001) completed a study in which they discovered that the results of their statistical analyses of gathered data changed significantly

when data were restricted to include only the participants that passed tests of effort. Consequently, failure to consider the validity of the client's performance and symptom reporting may result in inaccurate conclusions about the degree of impairment (Larrabee, 2012).

Defining Suboptimal Effort

Psychologists examining suboptimal effort are typically examining the client's motivation during the testing procedures, and the potential impact that varied motivation levels may have on the results of the assessments (British Psychological Society, 2009). When developing psychological measures, developers typically prescribe specific administration guidelines for the tests. These guidelines help ensure that subsequent administrations of the test provide the most valid and reliable results possible. Variations in administration may mean that the results are not reliable. As such, it is important that those taking psychological measures are not attempting to confound the results. Where optimal effort means that clients are administered and take tests consistent with developed guidelines, sub optimal effort indicates that clients are putting forth less motivation than is typical or ideal during administration of the test. Sub optimal effort is not a clinical diagnosis like malingering; however, research regarding neuropsychological measures typically uses the term malingering to mean a deliberate exaggeration of symptoms.

Incentives for Poor Effort in Different Environments

Forensic settings. Measures of effort are primarily administered in forensic settings (Bauer et al., 2007). This may be because exaggeration of symptoms is more abundant in forensic contexts than in other contexts (Bush et al., 2005). For example, a criminal defendant may have his or her sentence extended if it is determined that he or she did not put forth maximum effort on a court-ordered psychological evaluation (Kucharski, Ryan, Vogt, &

Goodloe, 1998). Additionally, defendants convicted of violent crimes are more likely to display traits consistent with Antisocial Personality Disorder, and also more likely to be found exerting suboptimal effort or malingering (Kucharski, Falkenbach, Egan, & Duncan, 2006). If a defendant can demonstrate an elevated pathology, it may lend credibility to certain defenses that excuse or justify his or her crime. Conversely, defendants may also deny pathology if they perceive that presenting themselves more favorably may decrease their served time. As a result, clinicians that complete assessments with the forensic system may encounter many individuals with incentive to complete psychological tests with poor effort (Bush, 2005).

Litigation. In personal injury litigation, pain and suffering may mean that claimants are entitled to damages. Because pain is subjective, litigants and their attorneys typically have a great incentive to exaggerate symptoms on psychological tests (Mendelson & Mendelson, 2004). The ability to demonstrate an enduring effect as the result of an accident or injury directly impacts the litigants' potential settlement or award. Research indicates that litigants with no head injury but with incentive to appear more cognitively injured than they truly are often score lower on tests of effort than individuals with moderate to severe head injuries (Suhr, Tranel, Wefel, & Barrash, 1997). In a study conducted by Schmand, Lindeboom, Schagan, Heijt, Koene, and Hamburger (1997), researchers discovered that participants involved in litigation underperformed on cognitive tests and demonstrated poorer rates of effort than patients not involved in litigation. Thus, exuding poor effort on neuropsychological tests is likely linked to substantial monetary gain for litigants.

Disability claimants. In a recent Congressional Response Report, government officials reevaluated the decision of the Social Security Administration (SSA) to disallow the use of symptom validity measures in the assessment of disability claims. The SSA sought to implement

the policy because no psychological test, by itself, could determine with certainty when clients where malingering or giving suboptimal effort. The committee assigned to evaluate the policy responded by stating that although such measures could not determine with perfect reliability those attempting to undeservingly gain benefits, an extensive amount of neuropsychological research has demonstrated that the use of such measures is critical in accurately assessing social security disability claims (Congressional Response Report, 2013).

Studies have indicated that as high as 30% of disability claimants have been determined to be giving suboptimal effort or malingering during psychological evaluation (Mittenberg, Aguila-Puentes, Patton, Canyock, & Heilbronner, 2002). Another study indicated that as many as 40% of claimants seeking disability for chronic pain disorders have been determined to be significantly exaggerating symptoms. Individuals with legitimate disabilities but who exaggerate symptoms, may be crying for help. Alternatively, individuals with mild to moderate symptoms may also exaggerate symptoms because they are aware that their claims are not likely to be approved ("Determinations of Malingering," 2005). Individuals attempting to claim government benefits may desire to do so in order to have a solution to socioeconomic problems. A consistent and predictable income, particularly if the claimant has some form of health concern that is not recognized as a disability, may feel justified. Disability claimants may also use the following as incentives to exaggerate their physical and psychological functioning: antisocial acts or behaviors; career dissatisfaction; work conflict; end of career; and also attempting to change a medical diagnosis to better fit the federal guidelines. Consequently, a great deal of disability claimants that exhibit suboptimal effort are responding to environmental or situational difficulties (Mittenberg et al., 2002).

College settings. Federal legislation requires colleges and universities to provide accommodations to students with diagnosed disabilities that interfere with the ability to function as a student (Gordon & Keiser, 1998). This is based, in part, on the principle of *equilibrium.* Referring to the American with Disabilities Act (ADA), equilibrium is the idea that students with diagnosed disabilities may need reasonable accommodations in order to perform at their true ability level (Gordon & Keiser, 1998). Such accommodations may include leniency with absences, private testing environments, access to note takers, and extended time for assignments (Lewandowski, Lambert, Lovett, Panahon, & Sytsma, 2014). Research has indicated that there is growing concern that college students may feign symptoms of disabilities, particularly of Attention-Deficit/Hyperactivity Disorder, in order to gain access to such accommodations (Jasinkski et al., 2011). In a study by Larrabee (2012), students with documented disabilities noted that having a separate room, a scribe, a reader, and word processor were of more benefit than did students without documented disabilities. Additionally, a significant number of students without disabilities stated that they believed that all students should have access to accommodations, or that tests should be redesigned so that accommodations are not needed by any student. This demonstrates that a great number of students in colleges and universities view academic accommodations as beneficial.

Research indicates that college students may give poor effort on psychological tests for a variety of reasons. Merckelbach and Merten (2012) posited that malingered symptoms may become internalized when the client experiences a great deal of cognitive dissonance. In other words, college students that believe their poor academic performance is the result of a disability may in turn begin to develop symptoms of that disability. Consistent with this finding, Suhr and Wei (2013) conducted a study in which college students used symptoms as an excuse for

performance. In this study, students in two different groups were asked to complete a task under two different premises – they are to play a computer game, or they are to complete a computer task that is a measure of intelligence. The task for both groups was identical, but one group performed under an evaluative threat. The study found that students that performed testing under the evaluative threat reported significantly higher ADHD symptoms, and were more likely to attribute poor performance to those symptoms. In a similar study conducted in South Africa, researchers found that students' motivation and effort exhibited on testing was the strongest predictor of academic ability (Goodman et al., 2011). These studies suggest that college students' attempting to show poor effort on tests in order to gain accommodations (external reward) may in fact be motivated by internal discomfort.

Studies report that as many as one-third of college students prescribed medication such as Adderall or Ritalin may divert that medication for one or more of the following purposes: recreational use, a source of income (resale), studying longer hours, or increasing concentration or ability to hyper-focus (Carroll, 2011). Additionally, with the number of job opportunities waxing and waning for college graduates, it may be the case that college students feel a great amount of pressure to be very successful. Students with other diagnoses such as depression or anxiety may feel the same pressures (Tan et al., 2002). As a result, they may perceive that accommodations will make them more competitive in the job market by increasing grade point average (Suhr & Wei, 2013).

Detecting Sub-Optimal Effort

A number of measures have been developed that are designed to assess whether or not a client is giving full effort on psychological tests (Rosenfeld, Sands, & Van Gorp, 2000). These tests may take one of two forms: (a) they may be stand-alone measures, or overt tests designed to

specifically detect effort, or (b) they may be embedded measures of effort (Babikian, Boone, Lu, & Arnold, 2006). Embedded measures are subtly integrated into a test that may have been designed for another purpose, such as assessing emotional difficulties. The decision to use stand-alone or embedded tests of effort may incorporate several dimensions (Schutte, Millis, Axelrod, & VanDyke, 2011). Stand-alone effort tests typically report higher psychometric face validity and accuracy, but may be susceptible to coaching influence (Victor et al., 2009). Because these tests are commonly used, it is easier for non-clinicians to research, study, and do well on the test. If a client can be coached to do well on a stand-alone effort test, but to do poorly on the psychological measures themselves, coaching may render the use of stand-alone measures useless. Alternatively, because clients may not perceive when embedded measures are assessing for consistency and exaggerated responses, they are often less susceptible to coaching. Embedded measures, however, may not be as accurate in distinguishing suboptimal effort from true psychopathology (Zeigler & Boone, 2013).

Because memory is the most commonly feigned impairment in neuropsychological evaluations (Constantinou, Bauer, Ashendorf, Fisher, & McCaffery, 2005), most stand-alone measures of effort focus upon detecting unnatural, unrealistic, or inconsistent patterns of memory (Schutte & Axlerod, 2013). Embedded measures of effort, however, may detect poor effort across multiple domains. Additionally, some research indicates that embedded measures may be more helpful than stand-alone measures if it becomes necessary to assess effort at multiple time points during the evaluation process (Schutte & Axlerod, 2013).

Stand-Alone Measures of Detection

Each measure of effort should be considered both in terms of its sensitivity and specificity. Sensitivity refers to the percentage of persons determined by the test to be exhibiting

poor effort, who are indeed putting forth poor effort. Specificity refers to the percentage of persons determined by the test to be exhibiting effort levels within normal limits, whom are actually giving adequate effort. Sensitivity identifies potential malingerers or suboptimal effort, whereas specificity identifies the population for which the neuropsychological data is likely an accurate reflection of their current level of functioning. Research indicates that effort tests, generally, have moderate sensitivity at 69% and high specificity at 90%. (Sollman, Ranseen, & Berry, 2011).

Test of Memory Malingering (TOMM). The Test of Memory Malingering is a neuropsychological assessment designed to distinguish between clients with true memory impairments and clients that are feigning memory impairments (Tombaugh, 1996). The test is designed to be able to screen for malingering via memory without also picking up other neurological impairments. The TOMM has two learning trials and an optional trial that assesses retention. The test uses visual stimuli, and signals potential malingerers by categorizing them as either below chance or by using criteria specific to clients with head injuries and cognitive impairments. The TOMM is considered the most widely used assessment of effort and malingering, and thus, has been the subject of countless validation studies in a variety of contexts with different populations. In a validation study with a non-clinical undergraduate sample and a sample of veterans, researchers found that the TOMM is particularly consistent in environments in which coaching is likely (Davis, Wall, & Whitney 2012). Other research, however, indicates that the TOMM may produce false positives up to 15% of the time. As a result, clinicians relying solely on the use of the TOMM to detect effort are advised to exercise caution in interpretation of the results. Several research studies have demonstrated that the TOMM is an appropriate screening tool for effort (Bauer et al., 2007).

Word Memory Test (WMT). The Word Memory test asks clients to memorize a list of twenty word pairs. Examples of pairs may include 'pencil-pen' or 'pig-bacon'. It is a stand-alone measure specifically designed to measure a person's effort on psychological testing. Clients are shown the twenty word pairs at the rate of one pair every six seconds. This procedure is repeated and the client is shown the list for a second time. Clients then are administered the Immediate Recognition (IR) trial (Green, Lees-Haley, & Allen, 2002). Results place clients in the categories of "Pass," "Caution," or "Fail." The Word Memory test has had extensive validation in clinical forensic settings. In a comprehensive study, researchers indicated that removal of data screened by the WMT and signaled as demonstrating poor effort, significantly affected the results of the study (Green, Lees-Haley, & Allen, 2002). Thus, the psychometric benefit of the WMT has been shown to be significant (Osmon, Plambeck, Klein, & Mano, 2006). A study by Greiffenstein, Greve, Bianchini, and Baker (2008) indicated that the WMT may be able to better distinguish effort levels when one trial of the TOMM is administered. Studies indicate that the WMT is psychometrically sound, boasting specificity as high as 99%. In other words, the WMT consistently identifies clients that put forth adequate effort (Gervais, Rohling, Green, & Ford, 2004).

Validity Indicator Profile (VIP). The VIP consists of 178 questions, 100 of which are non-verbal and 78 of which are verbal questions. The non-verbal questions consist of picture puzzles in which clients are to choose the puzzle piece that completes the picture. The verbal questions ask clients to choose the word most similar to the stem word given. Clients may not be told the name of the test, which may help decrease the likelihood of extreme symptom exaggeration (Allington, 2014). The Validity Indicator Profile examines the motivation and effort components of a client's test taking approach. A valid response style means the client

exhibited high motivation and high effort. A careless response style means that the client demonstrated some motivation but poor effort. A malingering response style indicates that the client had a high amount of motivation to appear impaired (Allington, 2012). One study regarding the specificity of the VIP indicated rates as high as 95%.

Embedded Measures of Effort

California Verbal Learning Test 2nd Edition (CVLT-II). The CVLT-II measures both recall and recognition over a number of trials. The trials encompass both immediate recall components and delayed recall components. Clients are asked to read a list of words and then recall words that they read. This occurs five times. They are then given a different list of words and asked to immediately recall them. The test then consists of short-delay free recall trial and a short-delay cued recall trial. Clients then complete non-verbal testing during a 20 minute delay. (Delis, Kramer, Kaplan, & Ober, 2000). After another series of long-delay recalls and a yes/no recognition trial, clients are then administered the Forced Choice Recognition trial.

The Forced Choice Recognition Trial of the CVLT-II may be used to detect suboptimal effort in individuals reporting memory and learning concerns and demonstrates strong predictive value in positive findings of inadequate effort (Root, Robbins, Chang, & Van Gorp, 2006).

Reliable Digit Span (RDS) of Wechsler Adult Intelligence Scale – (WAIS-IV). The Wechsler Adult Intelligence Scale is an assessment that aims to measure the intellectual capacity and ability of clients between the ages of 16 and 90. Individual subtests comprise indices helpful to clinicians. The indices on the WAIS include a verbal comprehension index, a perceptual reasoning index, a working memory index, a processing speed index, and a Full Scale IQ score. Individuals' scores are standardized by comparing raw scores to established norms from similar aged peers (Wechsler, 2008).

Reliable Digit Span is a subtest that has been used to assess client effort quite a bit (Schroeder & Marshall, 2011). In a study with 141 college students, researchers found that the RDS correctly identified the vast majority of students not exhibiting poor effort. The same study found that this subtest rarely produced false positive results, meaning that less than four percent of the time the test incorrectly identified a client as having poor effort (Harrison, Rosenblum, & Currie, 2010).

Integrated Visual and Auditory Continuous Performance Test-Advanced Edition (IVA AE). The IVA-AE is a test that measures response control and attention. It consists of three stages – the Warm-Up Period, the Practice Period, and the Main Test. Clients are asked to click the mouse only when they see a "3" or hear a "5." Clients are instructed to NOT click when they see a "5" and hear a "3." Thus, the test requires clients to exhibit sustained and focused attention. The test may also provide clinical information that can be used to better understand a client's concerns related to attention that result from other medical problems, such as head injuries and dementia. A study published in the Archives of Clinical Neuropsychology reported that the IVA-AE boasts high accuracy, adding that 81% of the scales could not be faked (Quinn, 2003). Clinicians can analyze specific combinations of scores in order to form the basis of a scale called the Malingering Analysis. The aforementioned study reports three equations that are indicative of poor effort on the IVA-AE: (a) when the auditory response control quotient is added to the auditory attention quotient and the result is less than or equal to 118 (ARCQ + AAQ ≤ 118); (b) when the visual response control quotient is added to the visual attention quotient and the result is less than or equal to 116 (VRCQ + VAQ ≤ 116); and (c) when the Full Scale Response Quotient is added to the full scale attention quotient and the result is less than or equal to 112 (FRCQ + FAQ ≤ 112).

Measures of Academic Ability

Wechsler Adult Intelligence Scale 4th Edition (WAIS IV) – Matrix Reasoning. Matrix Reasoning is an untimed subtest of the Perceptual Reasoning index. Perceptual Reasoning is intended to measure an individual's capacity to solve problems, organize thoughts, and examine rules and logical relationships. Matrix Reasoning asks clients to identify the missing picture in a matrix from five different options. The items on Matrix Reasoning assess visuospatial ability as well as simultaneous processing (Sobel, 2014).

Woodcock Johnson Test of Achievement 3rd Edition (WJ-III). The Woodcock Johnson test of Achievement tests reading, mathematics, written language, oral language, and academic knowledge. There are 22 subtests that assess these 5 domains. Different combinations of the subtests provide helpful interpretable clusters that help clinicians form a more comprehensive picture of an individual's cognitive and academic ability (Woodcock, McGrew, & Mather, 2001).

Letter-word identification. Letter-Word identification measures ability to identify words. Students are asked to read letters and words aloud from a list, without context (Mather, Wendling, & Woodcock, 2001).

Fluency scores. Academic Fluency is comprised of Reading Fluency, Math Fluency, and Writing Fluency, each of which is a timed subtest. Reading fluency assesses a student's ability to read simple sentences quickly. Students are given three minutes to read sentences and indicate whether they are true or false. Math fluency assesses a student's ability to perform simple addition, subtraction, and multiplication operations quickly. For Math Fluency, students are given a response book and asked to solve as many of the problems as possible. Writing Fluency

measures an individual's ability to formulate and write simple sentences quickly. For Writing Fluency, individuals are given three-word prompts and are asked to write as many sentences as possible in seven minutes (Mather, Wendling, & Woodcock, 2001).

Summary

Literature establishes that detection of effort level is critical in neuropsychological testing. Individuals may have a range of reasons to distort their current levels of psychological functioning, including monetary reasons, the possibility of decreased or increased jail time, and the ability to access accommodations in post-secondary environments that could potentially effect academic ability. While a great deal of research has examined the detection of poor effort regarding ADHD assessment, only a small body of research discusses whether or not embedded or stand-alone measures of effort impact academic ability. Thus, this study seeks to understand whether or not each type of measure of effort (embedded or stand-alone) can predict academic achievement, and to what degree.

References

Allington, J. P. (2014). *The psychometric properties of the validity indicator profile*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. (Order No. 3604665)

Assessment of Effort in Clinical Testing of Cognitive Functioning for Adults. (2009, October). Retrieved April 14, 2016, from http://www.bps.org.uk/sites/default/files/documents/assessment_of_effort_in_clinical_testing_of_cognitive_functioning_for_adults.pdf

Babikian, T., Boone, K., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*, 145-159.

Bauer, L., O'Bryant, S. E., Lynch, J. K., McCaffrey, R. J., & Fisher, J. M. (2007). Examining the test of memory malingering trial 1 and word memory test immediate recognition as screening tools for insufficient effort. *Assessment, 14*, 215-222.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., & National Academy of Neuropsychology Policy, & Planning Committee. (2005). Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology, 20*, 419-426. Carroll, L. (2011, May 17). ‘Steroids for school’: College students get hooked on ‘smart drugs’. Retrieved February 7, 2013, fromhttp://www.today.com/id/43050779/site/todayshow/ns/today-today_health/t/steroids-school-college-students-get-hooked-smart-drugs/#.URRNoJjnGzA

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J.M., McCaffrey, R.J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology*, 20, 191-198, http://dx.doi.org/10.1016/j.acn.2004.06.002.

Davis, J. J., Wall, J. R., & Whitney, K. A. (2012). Derivation and clinical validation of consistency indices on the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, *27*, 706-715. doi:10.1093/arclin/acs078 (http://www.sciencedirect.com/science/article/pii/S0887617704000824)

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *CVLT-II: California Verbal Learning Test Manual*. The Psychological Corporation.

Detection of Malingering in Disability Evaluations. (2005, December 1). Retrieved September 9, 2015, from http://primarypsychiatry.com/determination-of-malingering-in-disability-evaluations/

Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology*, *19*, 475-487. doi:10.1016/j.acn.2003.05.001

Greiffenstein, M. F., Greve, K. W., Bianchini, K. J., & Baker, W. J. (2008). Test of Memory Malingering and Word Memory Test: A new comparison of failure concordance rates. *Archives of Clinical Neuropsychology*, *23*, 801-807. doi:10.1016/j.acn.2008.07.005

Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., & ... Schlechter, A. (2011). An investigation of the relationship between students' motivation and academic performance as mediated by effort. *South African Journal of Psychology*, *41*, 373-385. doi:10.1177/008124631104100311

Gordon, M. & Keiser, S. (1998). *Accommodations in higher education under the Americans with Disabilities Act (ADA).* New York: Guilford Press.

Green, P., Lees-Haley, P. R., & Allen, L. I. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology*, *2*, 97-124. doi:10.1300/J151v02n03_05

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

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment*, *17*, 283-293. doi:10.1177/1073191109348590

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16*, 257-272. doi:10.1007/BF01044769

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

Jung, J. Y., & Gross, M. M. (2014). Highly gifted students. In J. A. Plucker, C. M. Callahan, J. A. Plucker, C. M. Callahan (Eds.), *Critical issues and practices in gifted education: What the research says (2nd ed.)* (pp. 305-314). Waco, TX, US: Prufrock Press.

Kucharski, L. T., Falkenbach, D. M., Egan, S. S., & Duncan, S. (2006). Antisocial personality disorder and the malingering of psychiatric disorder: A study of criminal defendants. *The International Journal of Forensic Mental Health*, *5*, 195-204. doi:10.1080/14999013.2006.10471243

Kucharski, L. T., Ryan, W., Vogt, J., & Goodloe, E. (1998). Clinical symptom presentation in suspected malingerers: An empirical investigation. *Journal of the American Academy of Psychiatry and the Law, 26*, 579-585.

Larrabee, G. J. (2012). Assessment of malingering. In G. J. Larrabee (Ed.), (pp. 116-159). New York, NY, US: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(4), 625-631.

Lewandowski, L., Lambert, T. L., Lovett, B. J., Panahon, C. J., & Sytsma, M. R. (2014). College students' preferences for test accommodations. *Canadian Journal of School Psychology, 29*, 116-126.

Mather, N., Wendling, B. J., & Woodcock, R. W. (2001). *Essentials of WJ III tests of achievement assessment*. Hoboken, NJ, US: John Wiley & Sons Inc.

Mendelson, G. & Mendelson, D. (2004) Malingering pain in the medicolegal context. *Clinical Journal of Pain, 20,* 423-432.

Merckelbach, H., & Merten, T. (2012). A note on cognitive dissonance and malingering. *The Clinical Neuropsychologist*, *26*, 1217-1229. doi:10.1080/13854046.2012.710252

Mittenberg, W., Aguila-Puentes, G., Patton, C., Canyock, E. M., & Heilbronner, R. L. (2002). Neuropsychological profiling of symptom exaggeration and malingering. *Journal of Forensic Neuropsychology, 3*, 227-240.

Neff, M. E. (2014). *Instrument selection and daubert admissibility in child custody evaluations: A survey of psychologists*. (Doctoral dissertation).Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. http://search.proquest.com/docview/1651620102?accountid=4488 (Order No. 0830404)

Osmon, D. C., Plambeck, E., Klein, L., & Mano, Q. (2006). The word reading test of effort in adult learning disability: A simulation study. *The Clinical Neuropsychologist, 20*, 315-324.

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18*, 379-395. doi:10.1016/S0887-6177(02)00150-6

Reynolds, C. R., & Horton, A. M. J. (2012). Basic psychometrics and test selection for an independent pediatric forensic neuropsychology evaluation. In E. M. S. Sherman, B. L. Brooks, E. M. S. Sherman & B. L. Brooks (Eds.), (pp. 41-65). New York, NY, US: Oxford University Press.

Root, J. C., Robbins, R. N., Chang, L., & Van Gorp, W. G. (2006). Detection of inadequate effort on the California Verbal Learning Test-second edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society, 12*(5), 688-696.

Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*, 349-359.

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist. 25,* 437-453.

Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, J. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*, 339-342. doi:10.1136/jnnp.64.3.339

Schutte, C., Millis, S., Axelrod, B., & VanDyke, S. (2011). Derivation of a composite measure of embedded symptom validity indices. *The Clinical Neuropsychologist, 25*, 454-462

Schutte, C., & Axelrod, B. N. (2013). Use of embedded cognitive symptom validity measures in mild traumatic brain injury cases. In D. A. Carone, S. S. Bush, D. A. Carone, S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 159-181). New York, NY, US: Springer Publishing Co

Sobel, N. K. (2014). *Detection of performance invalidity in neuropsychological assessments: Utility of the matrix reasoning subtest* Retrieved from ProQuest Dissertations & Theses Full Text; ProQuest Dissertations & Theses Global. Retrieved from http://search.proquest.com/docview/1459765810?accountid=4488 (Order No. 3598818).

Sollman, M. J., Ranseen, J. D., & Berry, D. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment*, *22*, 325-335. doi:10.1037/a0018857

Suhr, J., Tranel, D., Wefel, J., & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, *19*, 500-514. doi:10.1080/01688639708403740

Suhr, J., & Wei, C. (2013). Symptoms as an excuse: Attention deficit/hyperactivity disorder symptom reporting as an excuse for cognitive test performance in the context of evaluative threat. *Journal of Social and Clinical Psychology, 32*, 753-769.

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist, 16*, 495-505. doi:10.1076/clin.16.4.495.13909

Tombaugh, T. N. (1996). *TOMM. Test of Memory Malingering*. New York: Multi-Health Systems.

Urbina, S. (2014). Essentials of psychological testing (2nd ed.). Hoboken, NJ, US: John Wiley & Sons Inc.

U.S. Office of the Inspector General. (2013, September). Congressional Response Report on The Social Security Administration's Policy on Symptom Validity Tests in Determining Disability Claims. (Publication No. A-08-13-23094). Retrieved from https://oig.ssa.gov/sites/default/files/audit/full/pdf/A-08-13-23094.pdf

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23*, 297-313.