2009

Evalutaion of embedded malingering indices in a non-litigating, relief seeking sample: a partial cross-validation using control, clinical, and derived groups

Russell D. Pella
Louisiana State University and Agricultural and Mechanical College

EVALUATION OF EMBEDDED MALINGERING INDICES IN A NON-LITIGATING, RELIEF SEEKING SAMPLE: A PARTIAL CROSS-VALIDATION USING CONTROL, CLINCIAL, AND DERIVED GROUPS

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Russell Dominic Pella
B.S., University of Nebraska at Kearney, 1999
M.S., New Mexico Highlands University, 2004
August 2009

Table of Contents

LIST OF ABBREVIATIONS..... iii

LIST OF TABLES..... iv

ABSTRACT..... vi

INTRODUCTION AND REVIEW OF LITERATURE..... 1
Defining Malingering and Noncredible Neurocognitive Performance..... 4
Methods of Detecting Noncredible Neurocognitive Performance..... 7
The Use of Standardized Neuropsychological Measures in Malingering Detection..... 14
Embedded Validity Indicators..... 16
External Incentives in Educational Settings..... 48

PURPOSE AND RATIONALE..... 55
Questions and Hypotheses..... 55

METHODS..... 60
Participants and Data..... 60
Procedures..... 61
Measures..... 67

RESULTS..... 71
Hypothesis 1..... 71
Hypothesis 2..... 74
Hypothesis 3 & Classification of Noncredible Performance..... 78
Hypothesis 4..... 84
Hypothesis 5..... 86
Hypothesis 6..... 88
Hypothesis 7..... 89

DISCUSSION..... 94
Hypothesis 1..... 94
Hypothesis 2..... 96
Hypothesis 3 & Classification of Noncredible Performance..... 97
Hypothesis 4..... 101
Hypothesis 5..... 102
Hypothesis 6..... 102
Hypothesis 7..... 103
Limitations..... 106
Conclusions..... 110

REFERENCES..... 112

VITA..... 127

List of Abbreviations

ADHD	Attention-Deficit Hyperactivity Disorder
ANOVA	Analysis of Variance
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders
FSIQ	Full Scale Intelligence Quotient
IQ	Intelligence Quotient
LD	Learning Disabilities
MANOVA	Multiple Analysis of Variance
Max. Digits Fwd.	Maximum Digit Span Forward
M-STBI	Moderate to Severe Traumatic Brain Injury
MND	Malingered Neurocognitive Dysfunction
MTBI	Mild Traumatic Brain Injury
Neg. Impression Mgt.	Negative Impression Management
NOS	Not Otherwise Specified
PAI	Personality Assessment Inventory
PCS	Post-Concussion Syndrome
Reliable Digits	Reliable Digit Span
TBI	Traumatic Brain Injury
Vocabulary-Digits Span	Vocabulary Minus Digits Span
WMI	Working Memory Index
WAIS-III	Wechsler Adult Intelligence Scale – Third Edition
WAIS-R	Wechsler Adult Intelligence Scale – Revised
WMS-III	Wechsler Memory Scale – Third Edition

List of Tables

Table 1. Diagnostic Values for embedded Wechsler Adult Intelligence Scale (R & III) Digit Span Measures……..… 19

Table 2. Frequency and Percentage of Cases by Primary Diagnostic Category in the Clinical Group……. 62

Table 3. Frequency and Percentage of Cases by Co-Occurring Secondary Diagnostic Category in the Clinical Group……... 63

Table 4. Frequency and Percentage of Cases by Primary Diagnostic Category in the External Incentive Group…….…... 64

Table 5. Frequency and Percentage of Cases by Primary Diagnostic Category in the No External Incentive Group…….……. 65

Table 6. Cutoff Scores Corresponding to Index…….…… 70

Table 7. Univariate Analysis of Variance for WAIS-III Variables by Group (Clinical v. Control)…….…..…… 72

Table 8. Univariate Analysis of Variance for WMS-III Variables by Group (Clinical v. Control)……..…..….. 74

Table 9. Univariate Analysis of Variance for WAIS-III/WMS-III Validity Indices by Group (Clinical v. Control)……...…… 75

Table 10. Median, Means, and Standard Deviations for the Ord et al. Index by Group (Clinical v. Control)…….……..... 76

Table 11. Median, Means, and Standard Deviations for the Rarely Missed Index by Group (Clinical v. Control)……..... 77

Table 12. Univariate Analysis of Variance for PAI Validity Indices by Clinical Group (No External Incentive v. External Incentive)..... 78

Table 13. Participants Scoring beyond Cutoff Scores for Validity Indices by Group… 80

Table 14. Participants Scoring beyond Cutoff Scores for Validity Indices for Control and Total Clinical Group…….…... 81

Table 15. Number of WAIS-III/WMS-III Indices Failed according to Percentage of Participants beyond Cutoff Scores…… 82

Table 16. Frequency of Participants by Group Satisfying Criterion B from the Slick Criteria according to Index Type……..…..... 83

(list of tables continued)

Table 17. Rates of Noncredible Performance by History of Cognitive Problems…..… 85

Table 18. Frequency and Percentage of Primary Axis I Diagnoses in the Probable MND Group……..…. 87

Table 19. Frequency and Percentage of Cases by Co-Occurring Secondary Diagnostic Category in the Probable MND Group……... 88

Table 20. Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for Control Group……...…. 91

Table 21. Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for No External Incentive……..... 92

Table 22. Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for External Incentive - Not Probable Group……..…..…… 93

Abstract

Researchers have recently noted college students fail validity measures and base rate data are needed for students meeting Slick et al.'s criteria (1999) for malingering. The association between meeting Slick Criteria and subsequent recommendations (i.e., to receive external gain) is unknown as is the diagnostic utility of embedded validity indices in this population. The authors utilized archival data from: 1) a university psychological clinic (n = 986) and 2) a university student control sample (n = 182). Measures included the Wechsler Adult Intelligence Scale-III, Wechsler Memory Scale-III, and Personality Assessment Inventory. Empirically supported embedded validity indices were utilized to retrospectively identify suspected malingering patients. Group performance, according to level of symptom credibility and level of incentive seeking, was evaluated through a series of multivariate mean comparisons. Data are presented for frequency of falling in the noncredible range on all validity indices. Diagnostic statistics for each index are presented according to hypothetical base rates. Examination of receiving psychological recommendations to obtain external incentive (i.e., academic accommodations, medications, etc.) is reported according to incentive and credibility level. University patients explicitly seeking external gain, particularly those meeting criteria for malingering, demonstrated lower performance on the measures and received a higher rate of recommendations for academic accommodations and/or medications than patients not seeking external incentive. Nevertheless, a number of diagnostic statistics indicated some embedded validity indices lack specificity for malingering in university samples. The current study supports classifying patients according to level of incentive seeking when evaluating neurocognitive performance and feigned or exaggerated deficits.

Introduction and Review of Literature

Studies of response biases in psychology have evolved over time. Early clinical studies and applications largely addressed how self-report biases were manifested in an individual's psychological symptom presentation via symptom minimization or exaggeration (i.e., faking good, faking bad), and subsequently influence case conceptualization for treatment and healthcare utilization planning (Meehl & Hathaway, 1946). Thus, common terms such as symptom exaggeration or symptom magnification provided clinical insight into an individual's style of reporting symptoms. While still used in that manner, the study of response biases has expanded as the clinical practice of psychology has rapidly encroached into forensic endeavors within the past twenty years. In fact, Sweet, Peck, Abromowitz, and Etzweiler (2002) estimated that practicing clinical neuropsychologists spend an average of nearly 10 hours per week involved with forensic activities.

As psychological testimony regarding clinical cases is often proffered as legal evidence, its accuracy and validity have repeatedly been subjected to scrutiny. This has led to applying various legal precedents to expert testimony to establish admissibility and legal acceptability standards in court (i.e., *Daubert v. Merrell Dow Pharmaceuticals*, 1993; *Frye v. United States*, 1923; *Kumho Tire v. Carmichael*, 1999). One aspect of this matter entails whether or not a given individual who has undergone psychological evaluation has presented him- or herself in an honest fashion to the expert psychologist, and legal challenges frequently regard accurately interpreting psychological test data in the courtroom. This issue is particularly pressing in cases involving potential monetary and/or civil liberties implications.

Among other circumstances, the context of such expert testimony may include rendering opinions in simple civil litigation, complex disability determination, evaluating neuropsychological sequelae of traumatic brain injury (TBI) secondary to motor vehicle

accidents, or giving opinions regarding level of cognitive status following chemical exposure. Psychological testimony is often used by vested legal parties to ascribe legal responsibility for damages with the hope of obtaining compensation for those damages. In such cases there is usually at least one party (e.g., a claimant) alleging psychological injury in order to benefit from a secondary gain. Such a clear external incentive may serve as an influential factor in how an individual claimant presents psychologically, especially when monetary settlements or damage awards are proportional to the degree of injury that the claimant demonstrates. It then follows that persons undergoing psychological evaluation while seeking psychological and/or neuropsychological damage awards potentially have an interest in presenting with impairments. Consequently, there are significant implications for identifying, quantifying, and then accounting for negative response biases in psychological testing within the legal arena to recognize illegitimate claims (i.e., identify those with noncredible performance). On a societal level, noncredible neurocognitive performance has been shown to have substantial costs. It is estimated that feigned behavior is present in nearly 20% of all cases presenting for medical care each year, which results in nearly $5 billion for associated legal and medical expenses (Ford, 1983; Gouvier, Lees-Haley, & Hammer, 2003).

In their survey of clinical neuropsychology diplomates, Mittenberg et al. (2002) reported that out of 33, 531 patients, malingering prevalence ranged from 8% of medical cases to 35% of fibromyalgia/chronic fatigue cases, with high rates for disability, litigation, and criminal cases as well. Those percentages are an average across referral sources, and are particularly high in cases being referred from defense oriented sources in cases involving civil litigation. Moreover, Ardolf, Denney, and Houston (2007) recently reported that the combined rate of malingered neurocognitive dysfunction (MND) was 54.3% for a sample of serially referred criminal defendants. According to Mittenberg et al. (2002) the rate is much lower in patients not involved

with litigation or seeking compensation (approximately 7%). However, malingering prevalence rates vary widely across settings, and are influenced by factors including reason for referral and referral source (Bigler, 2006; Mittenberg et al., 2002). Not only that, but researchers have noted that neurocognitive complaints and symptom exaggeration (i.e., disrupted attention and concentration) occur in psychiatric groups, head injured individuals, and normal participants (Delis & Wetter, 2007; Gouvier, Cubic, Jones, & Brantley, 1992; Kirmayer & Sartorius, 2007; Lees-Haley, Earnest, & Dolezal-Wood, 1995).

In addition to exaggerating or fabricating symptoms for legal claims, there are multiple sources of external secondary gain (e.g., making insurance claims, avoiding military duty) that may prompt individuals to employ a negative response bias, which may affect the likelihood of malingering. However, malingering research most often pertains to legal settings, as the consequences for successful malingering under those circumstances are dramatic and typically involve multiple parties. Other areas of non-financial external secondary gain (e.g., medication-seeking, opportunity for service utilization) have gone relatively unstudied until recently (Henry, 2005; Osmon, Plambeck, Klein, & Mano, 2006).

Just as effort and financial incentives have been shown to account for significant variance in neuropsychological test performance (Binder & Rohling, 1996; Green, Rohling, Lees-Haley, & Allen, 2001), it is thought that other forms of incentive are also related to neuropsychological performance. Therefore, the current project seeks to explore the effects of non-financial external gain on popular neuropsychological tests in a young adult clinical sample with the expectation that performance in individuals overtly seeking non-financial incentives in the context of evaluations is poorer than in individuals not expressing potential for external gain. Moreover, it is likely the case that individuals expressing potential for non-financial external gain have a higher proportion of noncredible performance on test validity measures given symptom

exaggeration may be higher in such a population in order to obtain external gain. Thus, the application of existing methods for detecting noncredible neuropsychological performance needs to be extended to other populations without *financial* gain to explore effects of incentive level as it can be conceptualized on a continuum.

In this project, the author briefly describes a contemporary definition of malingering along with common and novel methods for detecting noncredible neurocognitive performance with particular emphasis on using traditional neuropsychological measures. Through that description, the author discusses the application of those methods to populations not involved in legal or other financial-seeking populations. Next, the author provides a discussion of literature weaknesses and strengths, and presents an overview of research of noncredible neurocognitive performance as it relates to sources of non-financial external gain in academic settings. Finally, a rationale for the current data analyses follows to address concerns raised by obvious gaps in the current literature. In particular, the author discusses identifying non-financial compensation seeking behavior in relation to noncredible neurocognitive performance and offers specific hypotheses for the project.

Defining Malingering and Noncredible Neurocognitive Performance

While the current *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR; APA, 2000) contains criteria for defining malingering as a v-code (not a diagnosis of a mental disorder *per se*), that description is considered by many to be inexact and vague. In response, researchers have offered other definitions and commentary to help guide clinical and forensic practice (Bianchini, Mathias, & Greve, 2001; Boone, 2007b; Greiffenstein, Baker, & Gola, 1994; Slick, Sherman, & Iverson, 1999). Among the formal definitions, Slick et al. (1999) have proposed the most cited and widely applied classification scheme of noncredible neurocognitive performance, referred to as the Slick Criteria. According to Slick et al. (1999), MND is "the

volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility." (p. 552). Their system includes three categories of malingering based on the degree of evidence and level of certainty that an individual is actually malingering. The categories range from *possible* to *probable* to *definite* malingering, with the *definite* category representing the highest degree of certainty that an individual is in fact malingering neurocognitive impairment.

In order to classify an individual according to one of those levels of malingering, that person must display specific behaviors falling along four distinct areas outlined by Slick et al. (1999). The first level of evidence (Criterion A) requires that the individual has a *substantial external incentive* to display a response bias at the time of evaluation. Criterion B specifies that within the context of a substantial incentive, there must be direct evidence of a response bias from objective neuropsychological testing. Criterion C requires the presence of a response bias from a self-report source (e.g., personality inventory). Finally, Criterion D stipulates that Criterion B and/or Criterion C data are not *fully* accounted for by psychiatric, neurological, or developmental factors.

In application, there are several combinations of Criterion B and C evidence that may be employed in order to classify an individual as a malingerer. For example, Slick et al. (1999) further subdivided Criterion B evidence (neuropsychological data) into six areas: 1) a *definite response bias* (i.e., below chance, $p < .05$), on one or more forced-choice tests, 2) a *probable response bias* represented by results from a "well-validated" test or index, 3) a discrepancy between current neuropsychological test results and objective profiles of known brain function and dysfunction, 4) discrepancy between test results and actual behaviors, 5) discrepancy between test results and informant reports, or 6) discrepancy between test results and the individual's known history. Therefore, meeting Criterion B is possible via six different ways.

Similarly, Criterion C is subdivided thusly: 1) self-reported history does not reflect documented history, 2) self-reported symptom endorsement is inconsistent with known patterns of brain functioning, 3) self-reported symptoms are inconsistent with current behavior, 4) self-reported symptoms are inconsistent with report from collateral contacts, or 5) there is evidence of symptoms exaggeration or fabrication on "well-validated" self-report measures.

The term *definite* MND refers to meeting A, B1 (below chance performance on neuropsychological testing), and D criteria. Meeting the classification criteria for *probable* malingering can be done in two ways. In both cases, the individual must have a *substantial external incentive* to appear impaired (Criterion A) and the findings from Criterion B and C are not secondary to known *bonafide* disorders (Criterion D). In the first case, a *probable* malingerer also meets two or more criteria from any one of B1-B6. In the second case, a *probable* malingerer also meets one criterion from neuropsychological testing (any of B1-B6) *and* one from self-report testing (any of C1-C5). The third, and least stigmatizing level of malingering, *possible*, is described as initially meeting Criterion A. Then, the individual must demonstrate at least any one of Criterion C level evidence *and* either the Criterion C behaviors meet Criterion D *or* criteria for *definite* or *probable* MND is met except that the behaviors in question may be partially explained by an actual disorder (Criterion D). While the diagnostic criteria is more liberal when applying a malingering label according to the Slick et al. (1999) criteria as one moves from *definite* to *possible* malingering, the level of certainty that an individual is really malingering may decrease accordingly. Additionally, if one does not meet Criterion A, but still satisfies all or part of Criteria B or C, the validity and credibility of performance would be called into question, but the person would not be labeled as a malingerer, but likely a factitious disorder instead.

Despite obtaining a high level of certainty of determining MND, undesirable and

unintended consequences are likely to occur as a result of labeling (Murphy, 1976; Thornicroft, Rose, Kassam, & Sartorius, 2007), especially when one is labeled a malingerer (Szasz, 1956). For instance, the dangers of mislabeling someone as a malingerer may include losing legitimate workers' compensation, disability, pensions, accommodations, needed healthcare services, or loss of liberty (Yelin, 1986). Furthermore, once that label has been applied, any future medical or legal claims made by the individual will likely be skeptically viewed. Consequently, using the term malingering in clinical settings is disfavored by many neuropsychologists (Slick, Tan, Strauss, & Hultsch, 2004). Rather, when clinicians are confronted with poor effort consistent with MND, they tend to indicate that test results are invalid, are inconsistent with the severity of the injury, that the results are indicative of exaggeration, or are noncredible (Boone, 2007b; Slick et al., 2004). In following that line of thinking against undue stigmatization, research and clinical practice recommendations in this field of study stress avoiding false positive findings of noncredible performance to avert subjecting clients to potentially harmful consequences (Bush et al., 2005).

Methods of Detecting Noncredible Neurocognitive Performance

As the study of malingering has evolved, so too have the methods of detecting invalid performance on psychological tests. There are several methods of evaluating malingered performance including behavioral observations, statistical techniques, qualitative analyses, and clinical records review among others. There are also specialized instruments to detect such performance. Specialized instruments usually represent techniques including symptom validity testing, performance curve, atypical performance on tests, floor effects methods, and validity indices. While those malingering-specific tests represent a unique class of instruments designed for a single purpose, there has also been an effort to detect noncredible performance indirectly, utilizing widely used neuropsychological measures.

Determining the usefulness of this indirect approach requires investigation of how the procedures perform in disparate samples across settings. Just as there is no single, sufficient method or technique for detecting noncredible cognitive performance in a given case, there is no uniformly agreed upon statistical or methodological procedure for establishing the usefulness of the various techniques. Rather, most researchers use multiple strategies (Ashendorf, O'Bryant, & McCaffrey, 2003; Bush et al., 2005; Iverson & Binder, 2000; Nies & Sweet, 1994) that have been established by how well they identify individuals from known groups of noncredible responders. In this manner, tests and measures used to identify noncredible responding are held to the same standards as measures used to document the claimant's impairments and disability. Therefore, classification and error rates of psychological procedures must also be reported to demonstrate test effectiveness (Gouvier, 1999, 2001; Gouvier, Hayes, & Smiroldo, 1998). The emphasis in this line of research is as much on the performance of the instruments as it is on the persons who complete them.

Consequently, evaluation of diagnostic prediction in this research area generally follows methods applied to the usefulness of medical tests that account for the sensitivity and specificity of tests (Baldessarini, Finklestein, & Arana, 1983). Whereas sensitivity refers to the ability of a measure to accurately detect true cases of a phenomenon, specificity is concerned with the ability of an instrument to accurately detect true non-cases. For instance, a test that correctly identifies all cases of a condition and does not identify any individual without the condition as having the condition, has perfect sensitivity (100% or 1.00) and specificity (100% or 1.00).

Accordingly, a test that deviates from perfect sensitivity and specificity results in incorrect classification of those who actually have the condition and those not having the condition. For illustration, if a test developed for early cancer detection had a sensitivity of .95 and specificity of .60, it would fail to detect cancer in 5% of those that actually have cancer

(false negatives) and would detect cancer in 40% of those who do not have the cancer (false positives). Therefore, a test with high sensitivity and low specificity would be sure to "catch" the condition early in most cases, but would also suggest cancer was present in those who actually do not have cancer, resulting in a test that "cries wolf" more often than it misses true cases. While such a scenario for a test may be desirable in some cases, such as making sure to identify life-threatening conditions early in the disease process, it may not be appropriate when there are notable negative consequences for those misidentified as having the conditions (i.e., tissue biopsies, quarantine, etc.). In the case of being labeled a malingerer, researchers have generally assumed that a 90% specificity level is appropriate to avoid undue stigmatization and other undesirable social and economic consequences (Babikian, Boone, Lu, & Arnold, 2006; Bianchini et al., 2001), but others consider a 10% false positive rate far too high (Langeluddecke & Lucas, 2004).

However, sensitivity and specificity are not the only values to consider when determining the usefulness of a test. Several commentators have emphasized that the *effectiveness* of a test also lies in its ability to improve diagnostic decision across varying base rates (Baldessarini et al., 1983; Bar-Hillel & Hogarth, 1990; Faust & Nurcombe, 1989; Gouvier, 1999, 2001; Gouvier et al., 1998) because the confidence of diagnostic decisions are dramatically influenced by the proportion of the population that presently has a condition. When sensitivity, specificity, and condition base rate are all considered in individual cases, decisions can reflect the degree of probability that a condition is present when a test indicates it is present (positive predictive values) and the probability of the condition not being present, as indicated by the test (negative predictive value). For example, in low base rate conditions, even tests with high specificity and sensitivity values may not improve accuracy over decisions grounded in base rates alone. Therefore, if sensitivity and specificity values have been established, test effectiveness, in terms

of positive and negative predictive values, can be calculated for a number of hypothetical base rate conditions. In this way, metrics are created that provide a level of probability indicating a given condition is or is not present depending upon a test finding. Therefore, research in the MND arena often contains values relating to diagnostic accuracy in order to demonstrate test effectiveness.

In the following section, the author briefly describes some of the specialized measures to detect malingering and then discusses empirically derived adaptation of traditional neuropsychological testing to that end. Throughout that discussion the reader is reminded that the phrase "well-validated" malingering instrument has a variety of meanings depending on the intentions and skepticism of the user (Green, 2007). Nevertheless, clinical usage of the procedures is informed by values supportive of diagnostic utility and effectiveness.

Symptom Validity Testing

Symptom Validity Testing is a methodology utilizing tests with forced-choice response formats (i.e., True/False, Multiple Choice) with responses representing unambiguously correct or incorrect answers. Since the early work of Hiscock and Hiscock (1989), adapted from Pankratz, Fausti, and Peed (1975) and Pankratz's (1979) procedures and methodology to identify conversion symptoms, symptom validity testing has held platinum status as the only certain way to identify MND (Bianchini et al., 2001). This set of procedures is founded on the statistical principle that an individual, given the opportunity to respond in a forced-choice format, should not respond incorrectly to more items than would be expected from random chance responding at the $p < .05$ level. If an individual does respond below the level predicted by chance, then it is thought that performance surely indicates a deliberate (or definite) attempt to respond incorrectly (Reynolds, 1998; Slick et al., 1999). However, below chance performance is rare in cases of known MND and most likely found in those making unsophisticated attempts at feigning. It is

known to be especially sensitive to coaching as a preventative intervention (for a discussion see Boone, 2007b; Youngjohn, 1995). Thus more sophisticated or idiosyncratic attempts at feigning likely go unidentified if relying solely on this type of testing.

Interpretation of symptom validity test performance can also be conducted to consider low performance that is not below chance levels, but is below performance of individuals with known neurocognitive dysfunction (Binder, 1993). Through this approach, it is possible to make inferences regarding clinically meaningful low performance by establishing optimal critical cutoff scores that differentiate those with established impairments and those displaying malingered/noncredible performance. By using those scores, tests can then be investigated according to their level of classification accuracy based on a number of factors. Such symptom validity tests include the Test of Memory Malingering (Tombaugh, 1997), Victoria Symptom Validity Test (Slick, Hopp, Strauss, & Spellacy, 1996), and Portland Digit Recognition Test (Binder, 1993).

Performance Curve, the Floor Effect, and Atypical Performance

Most neuropsychological and cognitive tests contain a collection of items and procedures that vary in degree of complexity and difficulty (Lezak, Howieson, Loring, Hannay, & Fischer, 2004). This is particularly true of standardized achievement and intellectual functioning measures (Wechsler, 1997a; Woodcock, McGrew, & Mather, 2001) as well as specialized neuropsychological instruments (Delis, Kaplan, & Kramer, 2001). Many of those tests tend to have easy items at the beginning of sets and conclude with more difficult items. Thus, the examiner can expect typical clients to miss progressively more items toward the latter portion of tests, rather than observing the opposite pattern. If the level of difficult items passed exceeds an expected level relative to easier items, then inferences can be made regarding the level of effort during testing (Rogers, 1997). Procedures using this rationale include the Dot Counting Test

(Rey, 1941), Validity Indicator Profile (Frederick, 1997), as well as qualitative analyses of performance (Johnstone & Cooke, 2003).

The easiest items of a neuropsychological test identify the "floor" or the most basic level of abilities measured by a test. Therefore, most individuals can be expected to complete them correctly at an exceptionally high rate. Similarly, tests of noncredible performance have been developed that appear difficult, but are actually quite simple with successful performance requiring a low level of cognitive skill. On such tasks (e.g., Rey 15-Item Test; Rey, 1941) even individuals who have sustained significant levels of acquired brain injury, experience pain, or are poorly educated perform well. If an individual does perform poorly on those items or tests, especially when compared to groups of neurologically compromised individuals, poor effort is suspected.

Related to floor effect and performance curve methods, the atypical performance methods rely on extant knowledge of brain functioning to identify noncredibility. One of the assumptions about memory functioning is that it is easier to recognize learned information than it is to retrieve learned information in a free-recall format (Binder, Villanueva, Howieson, & Moore, 1993). Given this general finding, it is unexpected that an individual's level of ability to recognize recently learned information is compromised relative to free-recall regardless of the level of neurological insult. Therefore, one would not expect to obtain results suggesting poorer recognition memory as opposed to recall memory. Moreover, recent investigations by Hilsabeck, LeCompte, Marks, and Grafman (2001) have capitalized on the findings that implicit memory remains largely intact even in amnesic patients. As a result, failure to learn and remember implicit information raises suspicion of suboptimal effort. Another method involves examining the differences between attention and memory performance under the assumption that attentional resources are required for learning and remembering information (Butters, Salmon, Cullum, &

Cairns, 1988). Therefore, poor attention relative to memory performance is also a marker for poor effort (Hilsabeck et al., 2003; Mittenberg, Azrin, Millsaps, & Heilbronner, 1993). Lastly, auditory attention span also remains relatively uncompromised by a host of neurological insults such as amnesia (Baddeley & Warrington, 1970), dementia (Carlesimo, Fadda, Lorusso, & Caltagirone, 1994), or pain (Wade & Hart, 2002), which also makes that area of functioning an attractive addition to malingering investigation (Greiffenstein et al., 1994).

Validity Indices

Conventionally, the use of validity indices has referred to a set of internal scales of a test for assessing the consistency, reliability, and honesty of performance. This type of approach is most often used in self-report personality tests or questionnaires to ensure that individuals have responded consistently to items with similar content (Morey, 1991; Tellegen et al., 2003). However, other types of invalid responses include endorsing more severe symptoms than those with demonstrated physical or psychological problems, denying common problems that most people admit to experiencing, or reporting extremely bizarre or infrequent symptoms suggestive of fabrication. It is assumed that the validity indices of self-report inventories add information regarding a respondent's response style, which can be utilized to make inferences regarding the manner in which they report psychological matters in general (i.e., under- or over-endorsement). For instance, if an individual endorsed a higher number of items on a scale developed to assess an exaggerated or negative impression (extreme or bizarre symptoms), then it is expected the individual responded to many other items in that manner, thus presenting a biased view of their current psychological status. The idea of including validity indices in neuropsychological tests has also received empirical attention (Boone, 2007b) as has the inclusion of indices on self-report measures to detect malingered cognitive status (Larrabee, 1998, 2003b).

The Use of Standardized Neuropsychological Measures in Malingering Detection

Neuropsychological instruments provide quantitative and qualitative data regarding particular domains of cognitive functioning (Mitrushina, Boone, Razani, & D'Elia, 2005; Strauss, Sherman, & Spreen, 2006). Most traditional neuropsychological measures are developed to primarily survey singular cognitive domains (Lezak et al., 2004). For example, the Rey Auditory Verbal Learning Test (Rey, 1964) was developed to assess learning and memory for orally presented words. In addition to unidimensionally focused tests, instruments containing multiple subtests assessing several cognitive domains, such as intelligence tests and executive functioning tests (Delis et al., 2001; Wechsler, 1997a), are typically administered to derive a profile of cognitive strengths and weaknesses across mental abilities. Patterns of performance on such tests have been examined for their use in detecting noncredible performance. Those investigations have usually been validated against criteria derived according to the principles mentioned above (i.e., atypical response patterns, deviation from known brain functioning, etc.).

There are several reasons for including validity indices derived from existing measures of neurocognitive ability. Using standard neuropsychological tests to glean information regarding individuals' effort level decreases administration time of testing batteries when administration of additional validity measures might be unduly burdensome. One reason that time is a limiting factor to consider in clinical practice is that administration of a neuropsychological battery often requires a significant commitment (several hours). This is a potential burden for both the clinician and client. The evaluation process can be extended substantially with the addition of standalone validity tests; thus, having "built in" validity measures can decrease the time of test administration. This is important because in order for results of psychological testing and neuropsychological procedures to be useful, they generally require patience, attention, cooperation, and sometimes endurance (Johnson, Lange, DeLuca, Korn, & Natelson, 1997;

Lezak et al., 2004) on the part of the client, all of which can be negatively affected by extended testing. For example, one widely used validity measure, the Portland Digit Recognition Test (Binder & Willis, 1991), requires roughly 45 minutes to complete in a forensic neuropsychological evaluation that typically requires over a day of testing.

Beyond the fact that testing can cause fatigue and may be aversive to the client, professional and market pressures require time efficient evaluations. For example, third party reimbursement for additional validity testing is unlikely (Piotrowski, Belter, & Keller, 1998), even for government mandated disability evaluations (such as in Louisiana), which makes it unlikely that validity tests are given in the majority of cases (Chafetz, Abrahams, & Kohlmaier, 2007). While approximately 79% of surveyed "expert" clinical neuropsychologists in North America commonly use at least one validity measure (Slick et al., 2004), the percentage used by other practicing psychologists is unknown, but is likely less. In fact, Sharland and Gfeller (2007) conducted a more recent survey of clinical neuropsychologists and reported that only 24.9% always, 30.7% often, and 28.6% sometimes include some type of validity measure in a neuropsychological evaluation. This is potentially problematic given the high rates of suspected noncredibility across settings, as testing effort and financial incentives have been shown to account for significant variance in neuropsychological test performance (Binder & Rohling, 1996; Green, Rohling, Lees-Haley, & Allen, 2001).

Moreover, validity markers within a given test can allow more sophisticated inferences made about performance on that test (Mathias, Greve, Bianchini, Houston, & Crouch, 2002). For instance, when conducting retrospective chart reviews where no specialized validity testing was administered, the reviewer can calculate and use derived validity index scores to infer effort level, which is often necessary in forensic cases. Despite the advantages of using embedded measures of validity, they remain among the least used validity detection methods by

neuropsychologists when making a decision of noncredibility and many neuropsychologists do not even know about them (Sharland & Gfeller, 2007). Regardless, developing indicators of noncredibility for popular multiscale intelligence and memory tests has received considerable research attention.

Embedded Validity Indicators

The Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997a), Wechsler Memory Scale – Third Edition (WMS-III; Wechsler, 1997b), and earlier versions of those tests (WAIS-R/WMS-R; Wechsler, 1981, 1987), are among the most popular adult intelligence and memory tests used by psychologists and neuropsychologists (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Lees-Haley, Smith, Williams, & Dunn, 1996). Additionally, the WAIS-III is the most commonly administered measure by clinical neuropsychologists with the WMS-III closely following as the second most used (Rabin, Barr, & Burton, 2005). Therefore, developing embedded detection methods and noncredibility indices in these two instruments has the potential to add meaningful information to routine evaluations given such wide popularity and test usage patterns.

Several attempts have been made to use aspects of those measures for detecting noncredible responses. Within that area of study, researchers have outlined how those embedded approaches may be used in order to meet requirements for specific level(s) of evidence (Criterion B) according to the Slick et al. (1999) criteria (Bianchini et al., 2001; Etherton, Bianchini, Heinly, & Greve, 2006; Larrabee, 2007a; Millis, 2004). Data regarding performance and classification accuracy of derived measures have been reported across multiple clinical, experimental, and normative samples.

Digit Span Measures

For ease of review, selected embedded digit span indices with corresponding diagnostic

statistics across research studies are presented in Table 1. As noted above, auditory attentional span remains relatively spared in individuals with a range of neurological problems (Wilde, Strauss, & Tulsky, 2004). Given that finding, suspicion of noncredibility is raised when individuals perform very poorly on measures of that ability. While there are several tests available to assess auditory attentional span via a digit repetition method (Benton, Sivan, Hamsher, Varney, & Spreen, 1994; Randolph, 1998), they all generally require participants to repeat increasingly longer strings of numbers that the examiner presents verbally. Examination of noncredible behavior using digit span tests has followed several lines of study as researchers have investigated a wide variety of particular types of performance including those based on age-corrected scaled scores, digit span forward, digit span backward, maximum digit span forward, maximum digits backward, time to complete single trials, total number of digits completed, digit span performance in relation to other subtest performance, and maximum length of consistently accurate performance, a.k.a. reliable digit span (Reliable Digits) (Babikian & Boone, 2007).

Since digit span performance is assessed by the WAIS-R, WAIS-III, and WMS-III, the majority of malingering research applications of digit span has been conducted using those measures. The administration, scoring, and response formats for those tests is very similar between the older and newer versions, and the current WAIS-III/WMS-III versions of the Digit Span subtests are identical to one another. Therefore, findings from early research are readily generalizable to those measures.

In this section, the author reports research regarding each of the well-studied digit span methods in turn. Throughout the literature, researchers' methodologies and sample characteristics vary from using unselected patients, patients displaying known noncredible performance, undergraduates who received instructions to malinger their performance, community control samples, and/or those with acquired brain injuries. Although some

researchers have examined many of the digit span indices in isolation, others have also taken a broader glance at how they perform together in terms of classification rates for MND.

Digit Span Age-Corrected Scaled Scores. In relation to college students instructed to perform their best, participants instructed to malinger have shown lower age-corrected scaled scores on the Wechsler Digit Span subtest (Bernard, 1990; Heaton, Smith, Lehman, & Vogt, 1978). Early classification studies of age-corrected scaled scores began with undergraduate analogue malingering research studies reported by Iverson and Franzen (Iverson, 1991; 1994). In their comparison of head injured and memory impaired patients (n = 28), simulating college students (n = 21), and simulating inmates (n = 35), a 1% false positive rate (misidentifying nonmalingerers as malingerers) was found for age-corrected scaled scores < 4 and a 2% false positive rate was found for scores < 5. Correct classification rates for malingerers were 82.5% and 90% respectively. Mean scores per group were as follows: simulating college students (M = 2.3, SD = 1.9), simulating inmates (M = 3.1, SD = 1.4), and the neurological patients (M = 8.3, SD = 2.3). Additional control groups with students (n = 21) and inmates (n = 38) resulted in normative performance for both groups (students, M = 10.7, SD = 1.8; inmates, M = 10.7, SD = 2.5). Iverson and Franzen (1996) also used a cutoff score of < 4 resulting in a 77.5% sensitivity and specificity of 100% for identifying simulated malingering in psychiatric patients (n = 20; M = 2.64, SD = 1.42) and simulating college students (n = 20; M = 2.8, SD = 2.59) as well as a control group of clinical noncompensation seeking neurological patients (n = 20; M = 8.15, SD = 2.83) instructed to give their best effort. Orey, Cragar, and Berry (2000) have also shown that simulating college students generally perform approximately one standard deviation below the normative standard (M = 7.92, SD = 2.92).

More recently, Vickery et al. (2004) assigned participants, from a sample of 46 individuals with demonstrated moderate to severe traumatic brain injury (M-STBI), to two

Table 1.

Diagnostic Values for embedded Wechsler Adult Intelligence Scale (R & III) Digit Span Measures

	Sample	Study Type	Cutoff	Accuracy
Digit Span Scale Score				
Iverson (1994)	College Students Inmates	Simulation	<4	99% Specificity
Iverson & Franzen (1996)	Psychiatric Patients	Simulation	<4	77.5% Sensitivity 100 Specificity
Iverson (1994)	College Students Inmates	Simulation	<5	98% Specificity
Iverson & Tulsky (2003)	ages 16-34	Standardization	<5	97.8% Specificity
Babikian et al. (2006)	Mixed Patients	Known Group	<5	>99% Specificity 32% Sensitivity
Greve et al. (2007)	Toxic Exposure	Known Group	<5	100% Specificity 22% Sensitivity
Iverson & Tulsky (2003)	Normative ages 16-34	Standardization	<6	95.3% Specificity
	Neurological	Standardization	<6	96.6%Specificity
Heinly et al. (2005)	Clinical Patients	Known Group	<6	93% Specificity 36% Sensitivity
	Elderly Stroke		<6	86% Specificity

(table continued)

(table continued)	Memory Disorder		<6	90% Specificity
	Psychiatric Patient		<6	93% Specificity
	Non-TBI Patient		<6	90% Specificity
Axelrod et al. (2006)	TBI; MTBI	Known Group	<6	97% Specificity 36% Sensitivity
Babikian et al. (2006)	Mixed Patients	Known Group	<6	93% Specificity 42% Sensitivity
Dean et al. (2007)	Mixed Patients	Normative	<6	81% Specificity
	IQ < 79 IQ = 50-69		<6	33% Specificity
Graue et al. (2007)	Community IQ <70	Simulation	<6	19% Specificity 68% Sensitivity
Greve et al. (2007)	Toxic Exposure	Known Group	<6	93% Specificity 46% Sensitivity
Trueblood & Schmidt (1993)	MTBI Patients	Known Group	<7	62.5% Sensitivity 99% Specificity
	Mixed Patients		<7	85% Specificity
Trueblood (1994)	Clinical Patients	Known Group	<7	86% Specificity 75% Sensitivity
Reliable Digits				
Heinly et al. (2005)	Elderly Stroke		<6	86% Specificity

(table continued)

Heinly et al. (2005)	Memory Disorder		<6	82% Specificity
Heinly et al. (2005)	Psychiatric Patient		<6	98% Specificity
Heinly et al. (2005)	Non-TBI Patient		<6	88% Specificity
Duncan & Ausborn (2002)	Patient	Known Group	<7	90.3% Specificity 56.6% Sensitivity
Larrabee (2003a)	M-STBI	Known Group	<7	100% Specificity 23.1% Sensitivity
Meyers & Volbrecht (2003)	Mixed Clinical	Known Group	<7	Unknown
Bianchini et al. (2005)	Clinical	Known Group	<7	93% Specificity 45% Sensitivity
Etherton, Bianchini, Greve et al. (2005)	Pain Patients, M-STBI	Known Group	<7	99% Specificity
Heinly et al. (2005)	Mixed TBI	Known Group	<7	96% Specificity 39% Sensitivity
Dean et al. (2007)	Mixed Patients IQ = 50-69	Normative	<7	33% Specificity
Dean et al. (2007)	Mixed Patients IQ = 70-79	Normative	<7	81% Specificity

(table continued)

Graue et al. (2007)	Community IQ <70	Simulation	<7	15% Specificity 56% Sensitivity
Graue et al. (2007)	Community IQ <70	Simulation	<7	15% Specificity 56% Sensitivity
Greiffenstein et al. (1994)	TBI	Known Group	<8	73% Specificity 70% Sensitivity
Greiffenstein et al. (1994)	PCS	Known Group	<8	89% Specificity 68% Sensitivity
Meyers & Volbrecht (1998)	MTBI	Known Group	<8	95% Specificity 77% Sensitivity
Duncan & Ausborn (2002)	Patient	Known Group	<8	71.6% Specificity 67.9% Sensitivity
Mathias et al. (2002)	TBI	Known Group	<8	93% Specificity 67% Sensitivity
Larrabee (2003a)	M-STBI	Known Group	<8	93.5% Specificity 50% Sensitivity
	MTBI, STBI, Neurological, Psychiatric	Known Group	<8	unknown
Bianchini et al. (2005)	Clinical	Known Group	<8	77% Specificity 62% Sensitivity
Bianchini et al. (2005)	Non-clinical Control	Known Group	<8	87% Specificity

(table continued)

Etherton, Bianchini, Ciota et al. (2005)	College Students	Simulation	<8	100% Specificity 65% Sensitivity
Etherton, Bianchini, Greve et al. (2005)	Pain Patients, M-STBI	Known Group	<8	92% Specificity 37% Sensitivity
Greve et al. (2007)	Clinical	Known Group	<8	89% Specificity 54% Sensitivity
Marshall & Happe (2007)	Community IQ <70	Normative	<8	31% Specificity
Etherton, Bianchini, Ciota et al. (2005)	College Students	Simulation	<10	80% Specificity 80% Sensitivity
Vocabulary-Digit Span				
Iverson & Tulsky (2003)	Normative	Standardization	>6	97.4% Specificity
Mittenberg et al. (2001)	Normative	Standardization	>5	89.9% Accuracy
Greve et al. (2003)	Patients	Known Group	>5	77-100% Specificity 0-50% Sensitivity
Mittenberg et al. (2001)	Normative	Standardization	>4	82.9% Accuracy
Mittenberg et al. (2001)	Patients	Known Group	>4	79% Accuracy
Axelrod & Rawlings (1999)	Cognitive Rehab @ 2Mo.	Normative	>3	70% Specificity
Mittenberg et al. (2001)	Patients	Known Group	>3	68% Accuracy
Mittenberg et al. (2001)	Normative	Standardization	>3	73.9% Accuracy

(table continued)

Mittenberg et al. (2001)	Patients	Known Group	>2	50% Accuracy
Iverson & Tulsky (2003)	ages 16-34	Standardization	>2	90% Specificity
Graue et al. (2007)	Community IQ <70	Simulation	>2	100% Specificity 0% Sensitivity
Marshall & Happe (2007)	Community IQ <70	Normative	>2	2% Specificity
Mittenberg et al. (1995)	Control, Litigants	Comparison	>1.53	70% Sensitivity
Iverson & Tulsky (2003)	TBI	Standardization	>1	87.4% Specificity
Iverson & Tulsky (2003)	Clinical	Standardization	>1	87.2% Specificity

groups. One M-STBI group was instructed to malinger on neuropsychological testing (n = 23) and one M-STBI group served as a control group (n = 23). Another sample of 46 community volunteers was matched to the M-STBI group on the basis of demographics and they were also assigned to either a simulation experimental group (n = 23) or a control group (n = 23). To increase the external validity of the experiment, the authors provided financial incentives for compliance with the instructional sets. In the context of other neuropsychological measures, the simulating M-STBI group had significantly lower age-corrected scaled scores (M = 7.7, SD = 2.3) than the control M-STBI group (M = 8.3, SD = 2.4). They also reported that a group of simulating community malingerers performed at similar low levels (M = 7.4, SD = 2.2) to the M-STBI malingerers, and well below community control volunteers (M = 10.9, SD = 3.2).

Other researchers have focused on reporting mean scores and classification accuracy rates in groups of clinical patients suspected of providing noncredible performance. In a preliminary study, Rawling and Brooks (1990) reported data on 16 severe cases of head injury contrasted with a "simulation" group consisting of 16 matched patients with mild traumatic brain injury (MTBI) suspected of feigning. Descriptive results from a full neuropsychological evaluation were provided. Based on those data, though no formal statistics were conducted, digit span appeared to be the lowest performed subtest from the WAIS-R that separated the groups with patients showing poor effort scoring lower than even the severely injured patients. While they described a follow-up validation study with additional patients, they did not report WAIS-R subtest information regarding those other patients.

Trueblood and Schmidt (1993) used a Digit Span age-corrected scaled score cutoff of < 7 when differentiating MTBI malingerers (n = 8; based on the Hiscock & Hiscock [1989] symptom validity test), a group of MTBI patients with questionable validity (n = 8), and two matched groups of nonmalingering MTBI patients (G_1 n = 8, G_2 n = 8). Their report of mean performance

is as follows: MTBI malingerers (M = 5.9), MTBI patients with questionable validity (M = 5.8), and the two MTBI control groups (M_{G1} = 8.9, M_{G2} = 10.8). In calculating operating characteristics from their study, sensitivity was 62.5%, and specificity remained high at 99% (one false positive). They also reported a 14.9% false positive rate (85% specificity) for an ill-defined mixed clinical group (n = 74), but no additional data were provided from that portion of their study. Trueblood (1994) later indicated that a cutoff of < 7 resulted in 86% specificity in two matched clinical control groups (n = 22; M_{G1} = 8.3, M_{G2} = 10.5). There was a 75% sensitivity for clients in a known clinical malingering group (n = 12, M = 4.8; based on the Hiscock & Hiscock [1989] symptom validity test). That score also identified 8 of 10 clients with questionably valid performance whose mean digit span scaled score was 5.6. One substantial limitation of Trueblood's (1994) study is that there was no information regarding neurological status any of the samples.

Two other studies reported group scaled score means for the Digit Span subtest. Youngjohn, Burrows, and Erdal (1995) reported that the Digit Span scaled score (M = 8.71, SD = 3.21) and Digit Symbol scaled score (M = 8.60, SD = 2.52) from subtests of the WAIS-R represented the lowest subtests scores for a group of MTBI patients with persisting symptoms seeking compensation (n = 55). No control or comparison groups were examined and no data regarding symptom validity tests or other validity measures were provided. Suhr, Tranel, Wefel, and Barrash (1997) reported digit span scaled score mean performance for a MTBI probable malingering group in litigation (n = 31; based on the Hiscock & Hiscock [1989] symptom validity test), a MTBI litigation control group (n = 30; nonmalingering), and a MTBI control group not in litigation (n = 20). Again, Digit Span performance was lowest in the malingering group (M = 6.3, SD = 2.4), but remained within normal limits for the litigating MTBI controls (M = 9.1, SD = 2.7) and nonlitigating MTBI controls (M = 9.0, SD = 3.0). They also included Digit

Span data from other small nonlitigating clinical control patients who endured severe head injury (n = 15, M = 9.5, SD = 2.8), were diagnosed with depression (n = 30; M = 8.7, SD = 2.9), or had significant somatic complaints (n = 29; M = 8.9, SD = 3.4). However, neither Youngjohn et al. (1995) nor Suhr et al. (1997) provided information regarding classification accuracy related to their Digit Span findings.

In more recent research, Heinly, Greve, Bianchini, Love, and Brennan (2005) reported mean values for digit span scaled scores in a mixed group of TBI patients with mild to severe injury. Slick et al.'s (1999) criteria (Slick Criteria) were applied to all patients utilizing the Test of Memory Malingering, Portland Digit Recognition Test, California Verbal Learning Test, unique responses to tests, and Minnesota Multiphasic Personality Inventory – Second Edition as evidence for MND. The groups consisted of patients with no incentive to malinger (n = 45, M = 10.48, SD = 3.76), non-MND patients with only incentive to malinger (n = 101; M = 9.74, SD = 2.88), suspected malingerers (n = 127; M = 8.40, SD = 2.60), probable malingerers (n = 53; M = 6.65, SD = 2.34), and those meeting Slick Criteria for definite MND (n = 12; M = 6.13, SD = 2.80). Sensitivity was low at 36%, but specificity 93% was excellent for Digit Span scaled scores $\leq$ 5 in comparing the entire TBI group with TBI patients determined to be malingerers. They also indicated that, in a 20% base rate condition, a score of $\leq$ 5 was associated with a 56% positive predictive power. Their findings also revealed that scaled scores of $\leq$ 5 were associated with more false positives in patients who were elderly, had endured a cerebrovascular accident (14% false positive rate), had substantiated severe memory disorders (10% false positive rate), were diagnosed with a psychiatric condition (7% false positive rate), or had a diagnosis other than TBI (10% false positive rate).

Babikian, Boone, Lu, and Arnold (2006) utilized a known groups design based on Slick Criteria to investigate the classification accuracy in a mixed group of litigating psychiatric and

neuropsychological patients with documented noncredible neuropsychological performance (n = 66). They used a combination of several validity measures to satisfy Slick Criteria including the Dot Counting Test, Rey Word Recognition Test, b test, Word Memory Test, Rey's 15-item test, and the Rey Auditory Verbal Learning Test. They also used data from a clinical neuropsychological control sample (n = 56) and nonclinical control sample (n = 32). The group with noncredible performance was the worst among the groups on the Digit Span subtest (M = 6.2, SD = 3.1) with the clinical (M = 8.8, SD = 2.8) and non-clinical controls (M = 9.42, SD = 2.20) scoring higher. Investigation of Digit Span subtest of ≤ 4 resulted in nearly perfect specificity, but the cutoff score resulted in poor sensitivity (32%). However, when raised to ≤ 5, specificity was within an acceptable range (93% for clinic patients), while raising sensitivity (42%) resulting in a positive predictive power of 51% in a 15% base rate condition.

Greve et al. (2007) examined Digit Span age-corrected scaled scores in a sample of litigating patients claiming injuries secondary to toxic chemical exposure. They further subdivided this group into a MND group (n = 46) based on Slick Criteria, an indeterminate group comprised of those failing only one validity index (n = 39), and a nonmalingering group (n = 38). They used multiple possible combinations across the Word Memory Test, Test of Memory Malingering, Portland Digit Recognition Test, Computerized Assessment of response Bias, California Verbal Learning Test, General Memory Index minus Attention/Concentration Index, Minnesota Multiphasic Personality Inventory–Second Edition, and Processing Speed Index to classify individuals according to the Slick Criteria. Based on that classification, they reported that the MND group (M = 6.56, SD = 2.71) scored lower than both the indeterminate group (M = 8.09, SD = 2.21) and the clinical toxic claimant controls (M = 9.08, SD = 1.86). While specificity of digit span age-corrected scaled scores ≤ 4 was 100%, sensitivity was poor (22%). When the cutoff score was raised to ≤ 5, specificity was held at an acceptable level (93%) and

sensitivity was raised to 46%, which resulted in a 59% positive predictive power at a 20% base rate. The level of Digit Span scaled score performance of ≤ 5 was associated with only a 36% sensitivity and 97% specificity for differentiating MTBI and TBI groups from a known malingering group (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006).

Beyond clinical and experimental studies of digit span age-corrected scaled scores, Iverson and Tulsky (2003) examined the WAIS-III/WMS-III standardization sample to investigate base rates of low performance in the general population. Delineated by age bands, they indicated that the cumulative percent of scaled scores ≤ 4 in 1,000 young persons (ages 16-34) was 2.2%, and remained low at 4.7% with a score of ≤ 5. Moreover, only 3.4% of those diagnosed with a significant neurological condition (n = 123; TBI, Korsakoff's syndrome, Alzheimer's disease, temporal lobectomy, chronic alcoholism) reported in the technical manual had a score ≤ 5 with an average score of 10.1 (SD = 2.8).

Despite few false positives in neurologically impaired groups, Graue et al. (2007) called into question using the digit span age-corrected scaled scores (≤ 5) procedure in individuals with impaired intellectual functioning (Fullscale Intelligence Quotient [FSIQ] < 70) by conducting a study of community volunteers. The first group served as a community control group with low average intellectual functioning (n = 10; FSIQ M = 80.2) and the second community group (n = 25; estimated FSIQ M = 82.1) was instructed to simulate mental retardation on a number of tests including the WAIS-III. The control group was recruited from a local agency providing services to individuals diagnosed with mental retardation (n = 26; FSIQ M = 60). Graue et al (2007) compared groups based on a number of derived validity indices from the WAIS-III, including age-corrected scaled score. Results indicated that the mental retardation simulation group (M = 4.6, SD = 1.8) performed similarly to those diagnosed with mental retardation (M = 4.5, SD = 1.7) for age-corrected scores. As a result, the sensitivity (68%) and specificity (19%)

were poor in discriminating the groups. These data support other recent work showing specificity (81%) for scaled scores ≤ 5 in a mixed group of patients with low intellectual functioning (n = 63, FSIQ < 79) who had no incentive to feign (Dean, Victor, Boone, & Arnold, 2007). If fact, Dean et al. (2007) reported a 33% specificity level for that cutoff score in individuals with FSIQ scores ranging from 50-69. Consequently, while research generally supports the use of scaled scores ≤ 5 as an acceptable cutoff score for identifying those giving poor effort, that score is not appropriate to use with those who have intellectual disabilities.

Reliable Digit Span (Reliable Digits). Greiffenstein et al. (1994) developed the Reliable Digits technique to optimally identify individuals with invalid performance. The authors described the technique as follows: addition of the longest string of repeated digits with no error over two trials in both forward and backward conditions. In their landmark study, the authors studied a group largely composed of litigating patients referred by insurance companies and attorneys claiming TBI deficits. They examined three groups: a TBI patient group (n = 33), persistent postconcussion group (n = 30), and a malingering persistent postconcussion group (n = 43). They identified the probable malingering group according to improbable performance on two or more measures, disability in their social role, contradicting collateral reports and symptom history, and remote memory loss. They reported that the malingering group performed significantly lower (M = 6.7, SD = 1.2) than both the persistent postconcussion control group (M = 8.9, SD = 1.1) and TBI control group (M = 8.8, SD = 1.2). They also reported choosing a cutoff score based on performance -1.3 standard deviations in the TBI group (≤ 7). Their cutoff resulted in 70% sensitivity and 73% specificity when comparing the malingering and TBI group, while that score had 68% sensitivity and 89% specificity when comparing the malingering group with the persistent postconcussion control group. Greiffenstein et al. (1994) also reported the base rate for malingering was 57% among the TBI patients, and a base rate of 59% among the persistent

postconcussion patients. Parenthetically, these participants were nearly all referred for independent medical evaluations from insurance company defense attorneys.

Meyers and Volbrecht (1998) studied a group of patients with MTBI (n = 47) involved in litigation with 49 referrals from treating physicians. They classified nine litigants as malingerers according to a forced choice procedure. Among the nine malingerers identified, seven (77%) were also identified by Reliable Digits $\leq$ 7, while only 5% of those with credible performance were misidentified as malingerers. In a later study, Meyers and Volbrecht (2003) incorporated Reliable Digits $\leq$ 6 into a decision tree among several other validity measures in a large clinical sample that included 32 normal controls. While they reported classification accuracy for their entire battery, they did not report any data specific to failure rates of Reliable Digits nor did they report levels of performance on that measure.

Larrabee (2003a) also incorporated Reliable Digits into a battery of multiple validity measures. He examined archival data from 24 closed head injured patients with definite MND (meeting Slick Criteria using the Portland Digit Recognition Test, below $p < .05$ level). He then compared that group to nonmalingering patients who suffered documented M-STBI (n = 27). Among other validity measures, Reliable Digits $\leq$ 7 was used to differentiate the known groups. Reliable Digits showed a sensitivity of 50% and specificity of 93.5%. None of the control M-STBI scored $\leq$ 6, while 23.1% of the patients in the known group were identified. He then cross-validated the findings to additional clinical groups of similarly classified known definite (n = 24; M = 7.37, SD = 1.92) and probable malingerers (n = 17; M = 6.82, SD = 1.67). They were compared to credible patients with severe TBI (M = 8.87, SD = 1.13), MTBI (M = 10.00, SD = 2.22), mixed neurological problems (M = 10.62, 2.34), and psychiatric diagnoses (M = 9.79, SD = 2.33). Those credible clinical groups were not significantly different from each other according to Reliable Digits scores. Despite reporting classification rates according to Reliable Digits from

the initial study, his follow-up cross-validation report lacked those data.

In their archival review of 151 workers' compensation TBI referrals, Mathias et al. (2002) administered at least one specialized malingering instrument (i.e., Test of Memory Malingering, Portland Digit Recognition Test) and classified each client according to the Slick Criteria. They further divided the patients into a control group with no incentive to feign and did not meet Slick Criteria (n = 30) and into a probable MND group (n = 24). A Reliable Digits $\leq$ 7 cutscore resulted in 67% sensitivity and 93% specificity with 68% positive predictive power and 91% negative predictive power in a 20% base rate condition. They also reported information about the false positive cases, and indicated both cases had severe TBI and borderline intellectual functioning (FSIQ < 80), suggesting increased false positives in persons with known neurological compromise.

Duncan and Ausborn (2002) applied Reliable Digits to case records of adult males (n = 187) in a forensic sample who had no evidence of neurological dysfunction. They retrospectively identified each case as malingering (n = 54) or not malingering (n = 134) based on self reports, interview, and the Rey 15-item test. They reported the malingering group (M = 5.81, SD = 3.40) performed lower than the credible group (M = 8.87 SD = 2.14) on Reliable Digits. Using Reliable Digits $\leq$ 6, sensitivity of identifying malingerers was 56.6% and specificity was 90.3%. When raising the cutoff score to $\leq$ 7, sensitivity was raised to 67.9%, but specificity lowered to 71.6%.

Etherton, Bianchini, Greve, and Heinly (2005) conducted a records review of 200 pain patients (most of whom had a financial incentive) referred by physicians, workers' compensation, and attorneys. The patients were assigned to a nonmalingering pain sample (n = 53) if there was objective medical evidence of physiological damage associated with pain and no evidence of self-report symptom exaggeration or poor performance on Test of Memory

Malingering, Portland Digit Recognition Test, or Minnesota Multiphasic Personality Inventory – Second Edition. The definite MND group consisted of patients meeting Slick Criteria. They also used an archival sample of M-STBI patients (n = 69) with no history of noncredible behavior. A large portion of the M-STBI and control pain group had some sort of external incentive. The credible pain group (M = 10.51, SD = 2.28) and control M-STBI group (M = 10.23, SD = 2.25) scored significantly higher than the malingering group (M = 7.20, SD = 2.95). They reported 37% sensitivity and 92% specificity with a 65% positive predictive power at a 20% base rate for Reliable Digits $\leq$ 7. They further reported that scores of $\leq$ 6 in M-STBI were associated with 99% specificity.

Heinly, Greve, Bianchini, Love, and Brennan (2005) reported Reliable Digits mean values for clinic data from their mixed group of TBI patients with mild to severe injury. Those patients with no incentive to feign (M = 9.91, SD = 2.30), non-MND patients with only incentive to malinger (M = 9.93, SD = 2.30), and suspected malingerers (M = 8.32, SD = 2.03) scored higher than probable malingerers (M = 7.15, SD = 2.02) and those meeting Slick Criteria for definite MND (M = 6.61, SD = 2.23). While sensitivity was low at 39%, specificity (96%) was excellent for Reliable Digits $\leq$ 6 in comparing the total TBI group with those TBI patients determined to be malingerers. They also indicated that, in a 20% base rate condition, a score of $\leq$ 6 was associated with a 71% positive predictive validity (p. 441). Using their classification scheme, across all credible TBI patients, none scored lower than 6 on Reliable Digits (100% specificity). However, they also reported that a score of $\leq$ 5 was associated with more false positives in patients who were elderly, had endured a cerebrovascular accident (14% false positives), had substantiated severe memory disorders (18% false positives), were diagnosed with a psychiatric condition (2% false positives), or had a diagnosis other than TBI (12% false positives). They also held that a score of $\leq$ 5 was associated with a 56% positive predictive

validity with holding specificity to 95% in a 20% base rate condition in the credible TBI sample. Based on their findings, in conjunction with additional literature, they recommended that Reliable Digits ≤ 7 is consistent with a negative response bias.

In another study Bianchini, Love, Greve, and Adams (2005) studied an archive of 11 patients referred for neuropsychological evaluation, 10 of who were involved in workers' compensation litigation. Nine of the patients met Slick Criteria for either probable or definite malingering. Using Reliable Digits ≤ 6, neither of the two credible patients were misclassified, while 5 of the nine malingerers were correctly identified. When using Reliable Digits ≤ 7, one (25%) of the credible was misclassified, and six of the nine (66%) malingerers were correctly identified.

Babikian et al. (2006) showed their noncredible performance group (based on Slick Criteria) scored low on Reliable Digits ($M = 6.7$, $SD = 2.4$) with the clinical ($M = 8.9$, $SD = 2.0$) and non-clinical controls ($M = 9.28$, $SD = 1.61$) scoring higher. Investigation of Reliable Digits of ≤ 7 resulted in high 62% sensitivity and poor specificity (77% for clinical patients, 87% for controls). When lowered to ≤ 6, sensitivity (45%) fell, but specificity (93%) was raised. This resulted in a 53% positive predictive power in a 15% base rate condition.

Axelrod, Fichtenberg, Millis, and Wertheimer (2006) included Reliable Digits among several other validity measures. Archival data were analyzed from patients with documented TBI ($n = 29$), patients involved in litigation who met Slick Criteria ($n = 36$), and a MTBI control group ($n = 22$). Presentation of means for the groups indicated the malingering group ($M = 6.3$, $SD = 2.1$) performed significantly lower than the TBI ($M = 8.5$, $SD = 2.1$) and MTBI ($M = 9.7$, $SD = 2.7$) groups. Their study did not include classification rates for Reliable Digits.

Greve et al. (2007) reported that those meeting Slick Criteria ($M = 7.37$, $SD = 2.25$) scored lower than both an indeterminate group ($M = 8.10$, $SD = 1.55$) and clinical controls ($M =$

9.47, *SD* = 1.74). Specificity of Reliable Digits ≤ 7 was 89% with a sensitivity of 54%, while lowering the cutoff to ≤ 6 altered the specificity to 97% and lowered sensitivity to 46%. However, positive predictive power was high (79%) at a 20% base rate. Thus, Greve et al. (2007) suggested that a score of ≤ 7 in the presence of objective brain pathology is an equivocal indicator of malingering.

In Graue et al.'s (2007) sample of feigned mental retardation, the community control group with low average intellectual functioning (*M* = 8.1, *SD* = 2.0) performed higher than the mental retardation simulation group (*M* = 5.9, *SD* = 1.7), which performed similarly to those diagnosed with mental retardation (*M* = 5.5, *SD* = 1.6). Classification accuracy between the simulated mental retardation group versus the community mental retardation group was poor using a cutoff of ≤ 6 (56% sensitivity, 15% specificity). Marshall and Happe (2007) also reported failure rates of Reliable Digits ≤ 7 in a sample of 100 individuals diagnosed with mental retardation (FSIQ *M* = 63, ranging from 51 to 74) having no obvious incentive to feign impairment. As a whole, the group scored below the Greiffenstein et al. (1994) cutscore (*M* = 5.8, *SD* = 1.7). Accordingly, they indicated a high false positive rate (69%) in that group, suggesting that this method of malingering detection is at best questionable when used with intellectually disabled clients. Dean et al. (2007) also reported high failure rates (33% specificity) of Reliable Digits ≤ 6 in patients with FSIQ ranging from 50-69 (in the intellectually disabled range). However, fewer patients with IQ's ranging from 70-79 were misidentified (81% specificity).

In addition to clinical studies, experimental designs have also been used to investigate the utility of Reliable Digits. Strauss et al. (2002) used an experimental design to assign 74 community recruited adults into one of three groups. The first group consisted of 27 individuals with a history of documented head injury (19 MTBI, 8 M-STBI), a naïve group (n = 26)

containing individuals without head injury history and had no professional contact with head injured individuals, an experienced group of patients with no history of head injury but did have professional contact with individuals who had experienced a head injury. Each participant was then assigned to one of two experimental conditions: 1) simulated malingering or 2) control condition. To investigate variability of performance across repeated administrations of Reliable Digits and other malingering instruments, all participants were tested on three occasions. They indicated that performance on Reliable Digits over testing trials was variable. They also replicated the finding that Reliable Digits performance was lower in experimental malingerers than controls, and this finding was robust regardless of the level of prior knowledge and experience with head injury. Mean Reliable Digits performance for the first testing session per group in the control condition ranged from 10.25 to 11.77 (*SD* = 2.93-2.20), while average malingered performance ranged from 5.60 to 6.46 (*SD* = 2.57-4.06). Classification accuracy for the use of Reliable Digits to correctly identify control participants was 90% and accuracy for the identified malingering participants was 79.4%. Thus, the sensitivity (47%) was poor, but specificity was high (95%) for Reliable Digits $\leq$ 6.

In another experimental study, Etherton, Bianchini, Ciota, and Greve (2005), assigned participants to one of three conditions: controls (*n* = 20), those instructed to simulate memory impairment (*n* = 20), and participants completed Reliable Digits while experiencing laboratory-induced pain (*n* = 20). Performance on Reliable Digits was significantly better for controls (*M* = 10.65, *SD* = 2.33) and in the laboratory pain condition (*M* = 10.25, *SD* = 1.62) than for the simulation group (*M* = 6.25, *SD* = 2.73). They reported that Reliable Digits $\leq$ 7 resulted in 65% sensitivity and 100% specificity and Reliable Digits $\leq$ 9 was associated with 80% sensitivity and specificity. The high specificity in this group likely reflects the high level of education of their university sample.

Digit Span Forward and Digit Span Backward. In the Digit Span task of the WAIS-III/WMS-III, the participant repeats the orally presented stimuli sequentially from the first number presented to the last. As the task progresses, each successive two-item trial is extended by a single digit. For example, the items are first trial of two digits in length, while the second trial is composed of items three digits long, and each trial increases sequentially. The total Digit Span forward equals the sum of the number of correctly completed trials. The test is discontinued when the participant misses both items of any given trial and the trial length before discontinuing represents the maximum digits span (forward). The methodology is the same for the backward span task, expect that the participant is required to repeat the digits in reverse sequential order. Thus, participants also obtain scores for the number of correct trials (Digit Span backward) and longest string of digits they are able to repeat (maximum attentional span backward). Research has consistently shown that participants instructed to malinger display a shortened attention span for repeating numbers sequentially (Iverson & Franzen, 1994, 1996; Vickery et al., 2004) and others have shown the same pattern in known noncredible patients (Babikian et al., 2006; Binder & Willis, 1991; Suhr et al., 1997). Iverson and Tulsky (2003) have reported normative performance of maximum attentional span forward and maximum attentional span backward performance and indicated that a maximum attentional span forward ≤ 4 and maximum attentional span backward ≤ 2 is rare in the general population. Heinly et al. (2005) and Babikian et al. (2006) also reported that maximum attentional span forward ≤ 4 rarely occurs except for those with known MND. However, while the occurrence for maximum attentional span backward ≤ 2 was rare in those shown not to be malingering, the rate was considerable among those with a documented cerebrovascular accident and memory problems. This later observation likely reflects the assertion that backward digit span tasks require working memory resources, which are often affected by those conditions (Reynolds, 1997). Despite the aforementioned

research with forward and backward spans, they are among the least studied measures in research order to establish optimal cutscores.

Difference Score Approaches

Researchers have also examined feigning behavior based on current knowledge of brain functioning. This has led to exploring improbable test patterns to identify malingering. For example, recognition memory is generally superior to free-recall memory. Based on that knowledge, a series of early studies addressed discrepancy of patients' performance on recognition versus recall (Bernard, McGrath, & Houston, 1993) and some have also examined discrepancy between attention/concentration and general memory performance (Hilsabeck et al., 2003; Iverson, Slick, & Franzen, 2000; Mittenberg et al., 1993; Slick, Hinkin, van Gorp, & Satz, 2001). Other discrepancies consist of examining domains of brain functioning that have been shown to remain robust following cerebral insult, such as crystallized verbal knowledge and digit span, with indices known to be vulnerable to disruption. This approach was first introduced by Wechsler (1955) in the hold-don't hold comparisons in the original WAIS.

Vocabulary Minus Digit Span. Mittenberg et al. (1995) first examined discrepancies between WAIS-R measures of verbal knowledge and Digit Span performance. This procedure involves subtracting the digit span age-corrected scaled scores from the scaled score of the Vocabulary subtest, which results in the Vocabulary minus Digit Span index. In the 1995 study, they compared 67 nonlitigating head injured patients to a group of adult volunteers instructed to malinger (n = 67) on the WAIS-R, and reported their findings in terms of Vocabulary minus Digits scores and associated malingering probability values. They showed that a Vocabulary minus Digits Span of 1.53 optimally differentiated the groups, which resulted in 70% sensitivity. However, they failed to report false positive rates for the control participants. Since that study, others have suggested using cutscores of two ($\geq$ +2) (Demakis, 2004; Millis, Ross, & Ricker,

1998). Larger discrepancies have been associated with higher levels of diagnostic accuracy in identifying noncredible performance, and WAIS-III Vocabulary minus Digits discrepancy scores where the base rate for malingering is 20% vary accordingly (Vocabulary minus Digits = 2, 50% accuracy; Vocabulary minus Digits = 3, 68% accuracy; Vocabulary minus Digits = 4, 79% accuracy) in differentiating malingered performance (Mittenberg et al., 2001). That finding was also echoed by Greve, Bianchini, Mathais, Houston, and Crouch's (2003) report that a Vocabulary minus Digits $\geq$ 4 was associated with the highest specificity (77% to 100%), but lowest sensitivity (0% to 50%) in relation to cutscores of two or three. Additionally, the rates of Vocabulary minus Digits $\geq$ 2 in clinical groups applying for vocational assistance or disability is near 30% (close to the hypothetical base rate; Williams & Carlin, 1999).

Despite the relatively good performance of Vocabulary minus Digits in differentiating malingering, there have been caveats issued by other research. For example, Axelrod and Rawlings (1999) studied archival data of three samples of cognitive rehabilitation patients with a documented significant TBI across multiple assessments throughout their recovery process. Using Vocabulary minus Digits $\geq$ 2, they noted specificities of 70% in all three groups at two months post injury. Specificity increased to 79%, 86%, and 91% in each group at a year status post injury. This suggests that performance on Vocabulary minus Digits, as a validity test, may not be appropriate for individuals in acute recovery stages of a TBI. Marshall and Happe (2007) also reported that Vocabulary minus Digits is likely insensitive to invalid performance in a mental retardation population as their sample of intellectually disabled patients had a mean score of -1.03 and only two of 100 patients scored below a discrepancy of greater than two. Graue et al. (2007) also reported that the average individual diagnosed with mental retardation has a negative Vocabulary minus Digits, while those malingering mental retardation only have slightly positive scores. This resulted in very poor sensitivity (0%) and perfect specificity. Taken

together, those results suggest that Vocabulary minus Digits in low IQ groups is not vulnerable to malingering. However, this may be due to the fact that those with lowered intellectual functioning tend to demonstrate overall lowered cognitive abilities. Thus, large discrepancies between cognitive abilities are less likely to occur in those with intellectual disabilities as performance is typically uniformly lowered.

Iverson and Tulsky (2003) further reported that difference scores (either positive or negative) of greater than or equal to five occur in fewer than 5% (2.5% for $\geq$ +5) of the WAIS-III standardization sample. They also found that difference scores of 2 or more occur in 13.6% of the TBI group reported in the manual and 6.9% of the total clinical patients. In young persons (ages 16-34), 20.9% have difference scores of $\geq$ 3, resulting in approximately 10% having positive values (Vocabulary minus Digits $\geq$ +3). They also reported that approximately half of the patients and clinical groups had a higher Digit Span score relative to Vocabulary. In general then, approximately 6.9% of patients in clinical groups have scores of $\geq$ 4, and 13.8% have values $\geq$ 3. Across all age groups, nearly 20.5% have Vocabulary minus Digits $\geq$ 3 (Vocabulary minus Digits $\geq$ +3). Moreover, Mittenberg et al. (2001) also reported the Vocabulary minus Digits of 2, 3, and 4 resulted in correctly classifying 73.9, 82.9, and 89.8% of the WAIS-III standardization sample respectively. Another report by Miller, Ryan, Carruthers, and Cluff (2004) indicated 99% specificity for Vocabulary minus Digits $\geq$ 6 in a mixed credible clinical sample. Therefore, a cutscore of $\geq$ 3 should result in adequate specificity in the general population and in most clinical settings where there is little evidence for marked neurological problems

Discriminant Functions

Mittenberg's Discriminant Function Score (Mittenberg Index). By applying multivariate techniques, Mittenberg, Theroux-Fichera, Zielinski, and Heilbronner (1995) differentiated

groups of simulated malingerers and nonlitigating patients according to performance pattern across WAIS-R subtests. Their procedure resulted in a widely used and cited classification discriminant function score (Mittenberg Index). They reported the following equation: Digit Span (-.3289) + Vocabulary (.1715) + Arithmetic (-.0720) + Comprehension (-.0811) + Similarities (.1580) + Picture Completion (-.0799) + Digit Symbol Coding (.0780) + .9696 (Constant). By using that combination of subtest scaled scores and derived coefficients, they reported that a Mittenberg Index value of > .00 resulted in 50% probability of malingering with the probability increasing as the value of the Mittenberg Index increases. Using the Mittenberg Index, they were able to obtain a 76% true positive rate. Further cross-validation work of the Mittenberg Index in a known malingering group and credible TBI sample showed a Mittenberg Index cutoff of .10536 resulted in high classification accuracy for both the TBI (92%) and malingering (88%) groups (Millis et al., 1998).

Their technique has also been further cross-validated on clinical groups using the WAIS-III. Mittenberg et al. (2001) reported a sensitivity of 72.2% (for simulating malingerers) and 83.3% specificity for a cutoff of > .00 as recommended by Mittenberg et al. (1995). The sensitivity for probable malingerers in a clinical sample was low (44.4%). Classification accuracy rates according to a 20% base rate across groups were reported for Mittenberg Index cutoff scores of .00 (51%), .10 (55%), and .20 (62%). Overall, their cross-validation supported use of the formula for extended application of the Mittenberg Index to the WAIS-III.

Greve et al. (2003) further applied the Mittenberg Index to classify malingering patients. However, their findings suggested that Mittenberg Index > .00 resulted in poor specificity (80%), but was increased when the cutoff was raised to > .212 (92% to 100% for the WAIS-III). That more conservative cutoff resulted in adequate sensitivity (57% for the WAIS-III). In general, they found a higher false positive rate in their S-MTBI group than the MTBI group, suggesting

the Mittenberg Index is somewhat sensitive to injury severity. However, no credible patient with MTBI was misclassified as a malingerer based on Mittenberg Index > .00, indicating that value may be appropriate for patients claiming less severe cerebral impairment (as in MTBI). Their calculations indicated that the positive predictive power for Mittenberg Index > .00 was 44% and negative predictive power was 88%. However, Bianchini et al. (2005) also applied the Mittenberg Index to their small sample of patients claiming neurocognitive dysfunction from electric injury. Inspection of their classification indicated a sensitivity of 67% and perfect specificity for a Mittenberg Index > .00.

However, there may be clinical situations where use of Mittenberg Index > .00 is inappropriate. Axelrod and Rawlings (1999) reported that the acute recovery phase of rehabilitation patients may be one such instance as specificity ranged from 76% to 78% at two months post TBI. Nevertheless, this high false positive rate declined substantially following a year after injury with specificity ranging from 88% to 93%. In their analysis of patients diagnosed with mental retardation, Graue et al. (2007) reported the average Mittenberg Index was .04 (*SD* = .74), thus sensitivity (48%) and specificity (65%) were not optimal. In all, positive findings from Mittenberg Index may be more commonly found in cases of documented severe impairment in those experiencing acute TBI recovery and in cases with *bonafide* developmental disabilities. Therefore, cutscores should be modified based on the clinical population (> .212 for S-MTBI, .00-.10 for MTBI patients and groups with no suspected neurological injury).

Rarely Missed Index. Poor recognition memory is often associated with profile invalidity (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005) and not significantly related to MTBI or M-STBI (Fisher, Ledbetter, Cohen, Marmor, & Tulsky, 2000). To capitalize on that observation, researchers have recently attempted to quantify the relationship between poor recognition performance and atypical performance on the WMS-III. After noting unusual

response bias patterns on the Logical Memory Delayed Recognition portion of the WMS-III (Killgore & DellaPietra, 2000a), Killgore and DellaPietra (2000b) further investigated items of that subtest that differentiated college students instructed to feign memory impairment from an archival review of credible neurological patients with memory impairment.

Using a discriminant function analysis, they determined that raw scores on the dichotomous items (yes or no) 12, 16, 18, 22, 24, and 29 were rarely failed ($p < .05$) in naïve control participants and patients. Their derived equation, the Rarely Missed Index, separated the groups with a high degree of accuracy. The resultant equation is as follows: item 12(-22) + item 16(55) + item 18(84) + item 22(67) + item 24(13) + item 29(7), and ranges from -22 to 226. Lower scores on the Rarely Missed Index were more associated with invalid performance and a cutoff score of $\leq$ 136 accurately identified 99% of the simulated malingerers. That score also resulted in a 100% specificity across base rates ranging from 1% to 50% with sensitivity remaining high (87%) in the lowest base rate condition.

Since that initial study, Langeluddecke and Lucas (2004) further supported the use of the Rarely Missed Index in an archival clinical sample of 99 consecutive referrals with MTBI. Almost half (46%) were plaintiff referrals and the rest were referrals from defense attorneys. Of that sample 28 patients met Slick Criteria and were compared to the remainder of the control sample according to the Rarely Missed Index. Given the groups' similarities (all were in litigation), the specificity (75%) of the Rarely Missed Index was poorer in their study with 100% specificity for cutoff scores $\leq$ 40. However, at that substantially lower cut-off point, sensitivity was very low.

In contrast, Miller, Ryan, Carruthers, and Cluff (2004) reported a 95% specificity for the original cutoff score in an archival sample of mixed, nonlitigating credible patients (n = 100) composed of various diagnostic groups (Alcohol Abuse, Polysubstance Abuse, and TBI). Two of

their false positive cases had been diagnosed with TBI had FSIQ scores of 81 and 91, and also had impaired memory performance. Another misidentified case also had low average memory performance. Two other false positive patients showed Vocabulary minus Digits ≥ 3. Thus, at least two of the false positive cases likely showed severe memory impairment, another had low memory functioning, and two others displayed questionable performance based on an external measure. This suggests that the participants misidentified by the Rarely Missed Index were actually questionable false positives. There is also recent evidence to suggest that the Rarely Missed Index may be appropriate to use in those diagnosed with mental retardation as it has only a 9% false positive rate in that population (Marshall & Happe, 2007).

Other Memory Indicators

Langeluddecke and Lucas (2003) demonstrated that low performance on the Auditory Recognition-Delayed subtest from the WMS-III was sensitive to malingering in an archival sample of TBI participants. They subdivided their sample into a credible ($n = 50$) and malingering ($n = 25$) group based on the Slick Criteria. They then selected the lowest score obtained by the credible group as the criterion for group comparison (Auditory Recognition-Delayed raw score < 43). This procedure resulted in 80% sensitivity and 91.8% specificity for the Auditory Recognition-Delayed cutoff score. Ord, Greve, and Bianchini (2007) also supported using cutoff scores based on the Auditory Recognition-Delayed standard scores ≤ 75. This cutoff also happens to correspond with Langeluddecke and Lucas (2003), with an Auditory Recognition-Delayed raw score < 43 for those ages 16-54 years (Wechsler, 1997b). They reported 94% specificity and 68% sensitivity with that indicator based on differentiating a known group of MTBI patients meeting Slick Criteria from those not meeting Slick Criteria. However, that specificity level dropped for patients with M-STBI and history of a cerebrovascular accident (89%) and dementia (73%). Conceptually, those findings are buttressed by findings that the

Auditory Recognition-Delayed subtest scores are not significantly lowered in MTBI and S-MTBI when compared to normal control participants (Fisher et al., 2000). Langeluddecke and Lucas (2003) further reported that the Logical Memory Delayed Recognition subtest (< 19) and Word Lists Delayed subtests (< 18) performed with similar specificity (91.7% and 96.6%, respectively), but differed in their sensitivity (48% and 81%, respectively).

Langeluddecke and Lucas (2003) also utilized the same method for the Faces I subtest (cutoff < 24), which resulted in 32% sensitivity and 96% specificity. This was supported by Glassmire et al.'s (2003) findings from a simulation experiment with nonmalingering M-STBI (n = 35) patients and participants with no history of neurological compromise (n = 30). Participants instructed to malingerer were compared on the Faces I subtest and found the cutoff of < 24 resulted in 100% specificity, but low sensitivity (33% to 63%).

In addition to reported cutoff scores for Auditory Recognition-Delayed, Ord et al. (2007) also published several classification rates of other standard index scores from the WMS-III. Those index scores associated with ≥ 90% specificity include the Auditory Immediate (≤ 80), Visual Immediate (≤ 80), Immediate Memory (≤ 75), Auditory Delayed (≤ 75), Visual Delayed (≤ 85), General Memory (≤ 80), and Working Memory (≤ 80) Indices. However, several of those cutscores are likely not appropriate for use with those diagnosed with neurological complications secondary to a cerebrovascular accident, dementia, or M-STBI as they result in a high false positive rate. To further investigate their classification system, they conducted a logistic regression using several of the indices as predictor variables and designating malingering status as the outcome variable. The resultant equation is as follows: Auditory Immediate(.0120) + Visual Immediate(-.0188) + Auditory Delayed(.0058) + Visual Delayed(-.0241) + Auditory Recognition-Delayed(.0001) + Working Memory(-.1169) + 10.890. From their equation, scores ≥ -1.0 have a 68% sensitivity and 91% specificity, with an associated likelihood ratio of 7.6,

whereas scores of ≥ -.5 are associated with 65% sensitivity and 97% specificity with a likelihood ratio of 21.7. Despite the well-performing equation, they recommended that clinicians identify noncredible performance by examining clients who score ≤ 75 points on three or more of those indices to avoid false positives. Thus, their approach appears quite robust and performed well in their sample. This is especially the case given their groups were based on the Slick Criteria and each group had a similar level of injury and most of credible patients also had incentive to perform poorly.

Working Memory and Processing Speed

Although cognitive processing speed and working memory are among the most sensitive brain functions to cerebral insult (Fisher et al., 2000), measures of those faculties have also been shown to be sensitive to noncredible performance. Several subtests from the WAIS-III/WMS-III assess those abilities including Digit Span, Letter-Number Sequencing, Arithmetic, Digit-Symbol Coding, Spatial Span, and Symbol Search. Across several studies of malingering employing both experimental and clinical malingering groups, researchers have reported significantly lower performance in malingering participants than nonmalingering participants. In particular, performance on Digit-Symbol Coding has repeatedly been low in malingering groups (Orey et al., 2000; Trueblood, 1994; Trueblood & Schmidt, 1993; Vickery et al., 2004; Youngjohn et al., 1995); however, that finding has not been universal (Rawling & Brooks, 1990).

In one of the earliest classification studies using thinking speed measures as markers for noncredibility, Trueblood (1994) reported a Digit-Symbol Coding scaled score of less than five (WAIS-R) was associated with 36% sensitivity and 100% specificity. Since then, others have reported low processing speed scores in patients displaying malingered pain behavior (Etherton, Bianchini, Heinly et al., 2006). Etherton et al. (2006) reported that those meeting Slick Criteria

for definite MND (M = 71.25, SD = 10.17) performed significantly lower on the WAIS-III Processing Speed Index than those with memory disorders (M = 83.24, SD = 12.82), M-STBI (M = 84.74, SD = 14.89), and control participants (M = 89.48, SD = 10.17). Although the noncredible group, as a whole, performed in the borderline impaired range of functioning, classification statistics from those data still revealed that a Processing Speed Index score of $\leq$ 70 had good sensitivity (63%) and borderline specificity (89%) in a mixed clinical sample (n = 121) containing TBI, pain, and patients with memory disorder. However, this approach to detecting invalid protocols is not likely indicated for use with those diagnosed with low cognitive functioning as in those with intellectual disabilities (Marshall & Happe, 2007). However, in samples with less significant cognitive dysfunction, that score would likely be appropriate and result in few false positives.

As reported above, the WMS-III Working Memory Index has also been shown to accurately classify those meeting Slick Criteria. Early reports of low WMS-III Working Memory Index scores among malingerers indicated that known probable malingerers scored nearly 18 standard score points (1.25 standard deviations) below control participants (Mathias et al., 2002). Lange et al. (2006) also reported that a clinical group meeting Slick Criteria performed nearly one standard deviation below a nonmalingering clinical control group on the WMS-III Working Memory Index. In accordance with this general finding, Ord et al. (2007) reported adequate classification rates for using the WMS-III Working Memory Index ($\leq$ 80) (94% specificity and 68% sensitivity). Moreover, Etherton, Bianchini, Ciota, Heinly, and Greve (2006) reported a WAIS-III Working Memory Index $\leq$ 70 resulted in specificity ranging from 93% to 96% with sensitivity of 47% in a sample of MND pain patients.

As a whole, the overwhelming focus on MND research has pertained to those seeking financial awards as the *substantial* incentive to perform poorly related to compensation for

alleged disability status caused by identifiable external factors. This is particularly the case concerning the use of the WAIS-III and WMS-III in detecting noncredible performance. Moreover, that research has largely pertained to comparing patients claiming significant brain injuries to patients with substantiated brain injury known to cause a number of functional impairments or comparison with patients determined to be malingering. In this regard, several of the aforementioned derived WAIS-III/WMS-III indices have been validated to discriminate patients exaggerating neurocognitive dysfunction.

However, feigning poor psychological test performance or exaggerating symptoms in order to establish disability status related to less severe neurocognitive dysfunction also occurs to obtain other external gains (not necessarily monetary) and pertains to those with alleged congenital neurocognitive problems. Nevertheless, very little research on noncredible test performance has been conducted with samples seeking other forms of external relief to accommodate or relieve disadvantages associated with neurocognitive problems less dramatic than TBI. For example, research into noncredible performance in those with disorders of isolated neurocognitive deficits such as specific learning disorders (LD) or Attention-Deficit Hyperactivity Disorder (ADHD) has been largely neglected. In such cases, there is no party responsible for neurocognitive deficits that are typically diagnosed early in the developmental lifespan. Hence, the problems evade the attention of the legal system as there are not typically large monetary sums at stake for obtaining associated disability status. Be this as it may, there are other tangible external incentives available as a function of being diagnosed with a less severe neurocognitive disorders such as ADHD or LD.

External Incentives in Educational Settings

There are numerous conditions (physical and mental), which might place students at a disadvantage when performing scholastically, resulting in high rates of poor academic

performance, grade retention, and dropping out (U.S. Department of Education, 2005). In order to address such disadvantages and potential discrimination against affected individuals within that class, governmental authorities have drafted legislation and adopted regulations to mitigate disabling functional effects of numerous health and/or social conditions. This is particularly relevant in the university setting as there has been a significant rise in the number of adults and post-secondary students complaining of cognitive problems and seeking disability status (Nichols, Harrison, McCloskey, & Weintraub, 2002). In fact, the number of individuals between the ages of three and 21 receiving support from federally funded educational programs for the disabled has nearly doubled within the past 20 years to over 6.7 million, and 1.7 million (8.7%) post-secondary students were considered disabled in 2000 (U.S. Department of Education, 2005).

While deficits negatively impacting scholastic performance can occur from functional limitations caused by a number of physical disabilities, they can also be attributed to a host of emotional, psychological, and/or cognitive disorders. For instance, those diagnosed with cognitive problems (e.g., ADHD, LD) often have functional deficits in thinking skills that negatively impact their ability to perform academic tasks. Therefore, educational institutions have formed policies and procedures in accordance with governmental statutes (e.g., Rehabilitation Act of 1973) to provide an accommodating environment for those with disabling conditions that result in specific functional deficits affecting academic performance.

A typical component of such procedures entails providing an academic environment tailored to individual student needs. This often includes providing the student with extra time on exams and assignments, taking tests in special settings (i.e., a quiet room), providing alternative response formats on exams (i.e., marking multiple choices on the test form rather than a Scantron response sheet), provision for note-takers, or offering other allowances (Evans, Serpell, Schultz,

& Pastor, 2007; Hadley, 2007). While such accommodations are available for eligible students through educational institutions, they are also offered to qualifying individuals taking important examinations that serve educational, organizational, and societal gait-keeping functions (such as the Scholastic Assessment Test, American College Test, Graduate Record Examination, or professional certification or qualifying exams). Due to the importance of such exams and the allure of academic success, accommodations in this regard may appear desirable to any student who has pressure to perform well, which provides an external incentive to seek disability status.

Another area of external incentive for those seeking to compensate for cognitive disabilities is obtaining medications that have been marketed to ameliorate cognitive deficits or improve cognitive functioning (i.e., amphetamines), especially for those with attentional dysfunction (Pary et al., 2002; Peterson, McDonagh, & Rongwei, 2008). In fact, prescription of amphetamine based psychostimulant medications (dextroamphetamine/racemic [Adderral], methylphenidate [Ritalin, Methylin, Concerta]) for attentional disorders (ADHD) has increased to high levels for students with documented attentional problems (McCabe, Teter, Boyd, & Guthrie, 2004). So too, non-prescription and non-medical adoption of such drugs have risen in normative student populations with increased levels of seeking and selling prescription medications for enhanced studying or recreational purposes. In fact, researchers have reported high rates of students either knowing someone who was prescribed a stimulant; or had personally taken and/or sold one themselves (Advokat, Guidry, & Martino, 2008; Babcock & Byme, 2000; McCabe, Teter, & Boyd, 2006; White, Becker-Blease, & Grace-Bishop, 2006). Moreover, thirty-four percent of college students prescribed a medication for ADHD reported being solicited for their medication (Moline & Frankenberger, 2001) and this trend has been satirized by popular media (Parker, 2000) as well. Therefore, university students may have compelling external reasons for seeking a disability status related to attentional problems in order

to obtain stimulant medication because they expect the use of stimulant medication will improve academic performance and for recreational purposes.

Thus, the potential for external gain secondary in being diagnosed with a disabling cognitive condition in the student population is a factor to consider when evaluating the nature and extent of attentional and learning complaints. This is particularly relevant as such disorders are often evaluated by psychologists (Crank & Deshler, 2001) employing measures with embedded indices shown to be sensitive to response bias and effort (see above). Despite this set of issues, there are currently no established measures available that have been validated to assess feigned deficits in ADHD assessment, and there is only one recently devised preliminary experimental measure to investigate effort in LD assessments (Osmon et al., 2006).

Of the few studies that have examined this matter in students seeking evaluation, most have employed simulation designs and/or have used small samples. For instance, Quinn (2003) showed that simulators of ADHD produced higher rates of attentional complaints and diagnostic symptoms of ADHD on self-report measures than controls or those diagnosed with ADHD. She also reported that the simulated ADHD group performed substantially lower on cognitive testing than both the control and clinical samples. Thus, her pioneering study revealed that test performance characteristics of ADHD can indeed be feigned through experimental manipulation. Other case reports in this area have also appeared in the literature and have suggested that malingered ADHD may be associated with behavioral problems (Conti, 2004).

More recently, Harrison, Edwards, and Parker (2007) examined the effects of instructions to feign ADHD symptoms and performance on self-report symptom inventories and measures of thinking speed. Similar to other simulation studies, they reported that their feigning group reported more symptoms than controls and a criterion clinical ADHD group. The feigning group also performed much lower on measures of thinking speed than the ADHD group and control

participants. Thus, instruments typically employed in an ADHD evaluation are sensitive to feigned poor performance and symptom exaggeration. Other simulation and clinical research has also shown that measures of thinking speed and attentional focus may be susceptible to feigned poor performance in typical neuropsychological evaluations (Henry, 2005; Leark, Dixon, Hoffman, & Huynh, 2002; Lu, Boone, Jimenez, & Razani, 2004; Willison & Tombaugh, 2006).

In the only published account of effort specific to LD assessment, Osmon et al. (2006) reported preliminary validation of a specialized assessment measure. In their simulation study, college participants ($N = 84$) were assigned to one of three conditions: 1) Reading Disorder simulation, 2) slow thinking speed simulation, and 3) normal effort. As part of the procedures, participants were administered the Word Memory Test (Green, 2003), the researchers' newly constructed Word Reading Test, and an estimation of intellectual functioning. The Word Reading Test is a computerized task developed to mimic lay persons' conception of reading disorders and dyslexic conditions. For instance, participants are briefly shown a word and subsequently shown two more words simultaneously – one of which is the initially presented target word (e.g., develop) and another word (e.g., bevelop) that may be attributable to reading problems (p. 316). Upon presentation of the two simultaneously presented items, the participant is required to choose the target item presented previously. Participants' performance on the task is assessed across multiple trials. Their rationale for devising this type of task was that individuals feigning reading problems will commonly err by choosing the non-word at a higher rate than those with diagnosed reading problems.

Their results supported that hypothesis, indicating appropriate usage of the Word Reading Test for detecting poor performance in reading disorder evaluations. They also reported that the Word Memory Test, a symptom validity test, classified simulating participants at an acceptable rate, indicating that traditional validity measures may be appropriate for this purpose. While the

Osmon et al. (2006) study represents the first and only published account of LD simulation, the results have yet to be replicated and extended to a clinical sample. Moreover, specialized measures, such as the Word Reading Test, are not available for clinical use, limiting the applicability of their findings. Relevant to LD assessment, though, poor Word Memory Performance was significantly associated with decreased performance on some memory and intelligence measures.

In a study of 67 consecutive referrals to a university-based psychological assessment clinic, Sullivan, May, and Galally (2007) investigated potential symptom exaggeration of students complaining of either ADHD and/or LD. They reported that 15 (22.4%) of the clients demonstrated failure on at least one index from the Word Memory Test, a validated measure of effortful responding. They also partially replicated inverse relationships between Word Memory Test performance and intellectual functioning in those failing the Word Memory Test. An inverse relationship was also demonstrated in clients who failed the Word Memory Test and the California Verbal Learning Test–Second Edition.

Although that study demonstrated a high failure rate on a symptom validity test and showed small to moderate positive relationships with objective cognitive measures, the authors may not have adequately addressed a few issues in the sample. First, they implied that the clients had external incentive for seeking the evaluation, but did not indicate this through formal means, simply indicating that the potential for secondary gains (i.e., accommodation-seeking or medication-seeking) in the sample was "operant" (B. Sullivan, personal communication, April 24, 2008). They also did not address the presence of pre-existing learning problems or attentional difficulties due to the lack of criterion control and questionable integrity of prior diagnoses. Moreover, the relationship between Word Memory Test performance and the cognitive measures was based solely on clients scoring below cutscores indicative of invalid performance, possibly

inflating the relationship. However, the authors only examined composite IQ scores, which likely masked the relationships of Word Memory performance and specific intellectual abilities (e.g., cognitive processing speed). Yet another weakness in the study is that only one measure of validity was evaluated, limiting the applicability of the findings to other established measures, and especially embedded indices. Lastly, the study lacked a control sample against which to compare the elevated failure rates of the clinical sample.

Purpose and Rationale

Reports of high failure rates on validity indices has only very recently been noted in college samples seeking psychoeducational evaluation but standard research or clinical criteria for noncredible performance (i.e., Slick Criteria) have yet to be applied in such samples. While Sullivan et al. (2007) reported evidence of high failure rates in this population, their lack of reporting potential for explicit external gains limits the application of the Slick Criteria to their sample. Thus, reasons for the lowered performance could be due to factors other than malingering as outlined in the Slick Criteria (Delis & Wetter, 2007). As such, the question of overt medication-seeking and/or academic accommodation seeking status has not been investigated for its possible relationship with performance on neuropsychological tests, nor to validity measures embedded therein. The focus of this project was to address some of those concerns.

Validity indices for the WAIS-III/WMS-III have received considerable attention and have been examined in simulated malingering groups, various patient groups (litigating, TBI, pain, etc.), and control participants. Nevertheless, the aforementioned validation studies often have employed small to moderate sample sizes and none have attempted to investigate simultaneously several validity indices from both the WAIS-III and WMS-III in the same study, even in small samples. Additionally, a thorough literature review failed to uncover published work reporting the wide range of WAIS-III/WMS-III indices in a university sample. This is a serious concern given the WAIS-III and WMS-III are the neuropsychological instruments most likely to have influence, not only into the clinic and courtroom, but in the academic setting as well.

Questions and Hypotheses

The purpose of this dissertation was to investigate the following questions:

Question 1

Is overt academic accommodation and/or stimulant medication seeking associated with performance on intelligence and memory tests?

Hypothesis 1. In keeping with findings that performance on neuropsychological tests is often feigned in those seeking other forms of external gain, it is expected that those overtly seeking recommendations for academic accommodations and/or stimulant medications (External Incentive) will show significantly lower performance on those measures when compared to a normal control group (Control) and a clinical group comprised of clients not overtly expressing interest in receiving such external incentives (No External Incentive). More specially, the author expects to find particularly lower performance on measures requiring thinking speed and attention (e.g., Digit Span, Letter-Number Sequencing, Symbol Search, Digit-Symbol Coding) in the External Incentive group as those symptoms have been shown to be among the most sensitive to noncredible performance.

Question 2

Do those seeking psychoeducational evaluations for overt external incentives differ from those not overly seeking external incentives and control participants according to validity indices from the WAIS-III/WMS-III?

Hypothesis 2. Just as litigation status has been a variable to consider in forensic contexts because of its association with poor neuropsychological performance, it is hypothesized that overtly seeking accommodations and/or stimulant medications will be associated with inflated failure rates on tests of noncredible performance. Therefore, it is expected that the External Incentive group will perform significantly more poorly from the other two groups (No External Incentive and Control) on all validity indices. For those indices where low performance indicates a high probability of noncredible performance, it is expected that the External Incentive group

will show the lowest mean scores. For those indices where higher scores indicate an increased probability of malingering, it is expected that the External Incentive group will show the highest mean scores. Moreover, the author expects that a higher proportion of clients in the External Incentive group will score in the noncredible range on the validity indices than those in the other groups.

Question 3

Does the proportion of positive cases and overall proportional degree of severity of malingering vary positively along the control – no external incentive – external incentive dimension?

Hypothesis 3. Given that the clinical setting of this study is utilized by a high number of those seeking external gain, and that recent research has demonstrated a high failure rate on the Word Memory Test in a similar population, the rate of noncredible performance according to the Slick Criteria in the entire clinical sample is expected to be similar to rates in other general clinical settings where compensation seeking is a factor (20%). That rate should skew highest among persons in the External Incentive group as rates have been shown to be higher in groups containing an increased number of clients overtly seeking external gain (Ardolf et al., 2007; Mittenberg et al., 2002). Moreover, the rate of noncredible performance is expected to progressively increase according to group along the incentive seeking dimension (control, no external incentive, external incentive).

Question 4

Do those reporting a history of previous psychological diagnoses have a higher rate of noncredible performance than those who did not report such a history?

Hypothesis 4. It is hypothesized that those with self-reported, pre-existing cognitive disorders will have a lower rate of noncredible performance. It is thought that those individuals

are less likely to exaggerate their deficits in order to obtain external gain, since the basis for obtaining accommodations has been established by previous health providers – often years prior to the current evaluation. On the other hand, those seeking new diagnoses may have a higher need to overstate or exaggerate deficits since they have demonstrated satisfactory scholastic performance to the degree that qualified them for university study.

Questions 5 and 6

What is the association between meeting Slick Criteria for probable MND and the outcomes of ultimate diagnosis and treatment recommendations? Is there an association between being categorized as putting forth noncredible testing behavior, according to the Slick Criteria, and likelihood of obtaining accommodations and/or medication referrals as recommended by the psychological clinic?

Hypothesis 5. Preliminary findings from simulation studies (Harrison et al., 2007; Osmon et al., 2006) and initial clinical observations of failure of indices from the Word Memory Test have been associated with decreased performance on some cognitive measures in LD and ADHD malingering. Moreover, effort has been associated with a considerable amount of variance in cognitive testing performance (Green et al., 2001). Accordingly, it is expected that a disproportionate percentage of those meeting Slick Criteria in the current study will be diagnosed with current brain based disorders (e.g., Cognitive Disorder NOS, ADHD, LD) as an outcome of their current psychoeducational evaluation.

Hypothesis 6. Because it is thought that those meeting the Slick Criteria for probable MND exaggerate neurocognitive deficits on objective measures, they are also more likely to receive recommendations to address those apparent deficiencies if their simulation goes undetected. Moreover, since most clinic evaluations did not utilize validity indices and that researchers claim behavioral observations alone are poor indicators of noncredible performance

(Faust, Guilmette et al., 1988; Faust, Hart, & Guilmette, 1988; Frederick, Sarfaty, Johnston, & Powel, 1994; Heaton et al., 1978), those retrospectively meeting Slick Criteria as per the current study, were likely not identified as putting forth malingered performance at the time of their evaluation. Therefore, those clients meeting Slick Criteria were probably most likely to be recommended to receive academic accommodations and/or recommended for a medication consultation at a higher rate than those not meeting criteria for probable MND because they demonstrated impairment on cognitive testing sufficient to justify obtaining a diagnosis and accompanying recommendations.

Question 7

What are the specificity values for each derived WAIS-III/WMS-III validity index when compared with a nonclinical control sample and a clinical sample expressing no overt external incentive?

Hypothesis 7. Specificity levels of many of the WAIS-III/WMS-III embedded indices have shown to be lower when groups meeting Slick Criteria are compared with non-MND clinical groups as opposed to comparison with normal control participants. The specificity levels of each validity index in this study are expected to be greater when those meeting the Slick Criteria are compared to the nonclinical control group rather than when compared to those not overtly seeking external incentives those seeking external incentives but not meeting Slick Criteria for probable MND.

Question 8

How effective is each test according to varying theoretical base rates when considering sensitivity, specificity, and hypothetical base rates of noncredible performance? As this is an entirely exploratory matter, no specific hypotheses are forwarded for this research question.

Methods

Participants and Data

The author utilized archival data from two sources for this project. The first dataset consisted of normative data from a portion of a large-scale joint experimental and clinical working memory project conducted between May 2005 and December 2006 at a public university in the southern United States. As part of that construct validation research protocol, participants were administered tests of intellectual functioning, memory functioning, and experimental measures of cognitive functioning. For the purposes of this study, the author utilized only intellectual and memory functioning data forming a large control sample. Of the 224 participants that originally agreed to participate in that study to obtain extra credit in undergraduate psychology courses, complete data from 182 participants were obtained. Exclusion criteria included visual and/or hearing impairment (n = 14), a psychiatric diagnosis resulting in cognitive impairment (n = 0), English as a second language (n = 5), or if the participant did not complete the study protocol (n = 23). No participant was excluded according to gender, race, ethnicity, or academic status. All participants were provided informed consent as part of the research protocol. Mean age in the control sample was 20.56 (SD = 3.67) with 47 (25.8%) males and 135 females (74.2%). Most participants were Caucasian (149, 81.9%). Another 15 (8.2%) were African American, 6 were Hispanic (3.3%), and 12 (6.6%) were identified as other.

The second data source consisted of information obtained from clients seeking services at the university's Psychological Services Center from 1999 to 2008. Exclusion criteria included visual and/or hearing impairment, current psychiatric or neurological disorders warranting legal disability status, history of moderate-severe traumatic brain injury, stroke, dementia, chronic/severe neurological condition(s), or English as a second language. While the final

clinical sample was 986, the clients were administered a fixed/flexible battery and therefore the individual test-by-test analyses presented here may not employ the entire sample but be limited to those who completed the particular measure of interest. Participants in this study were typically individuals self-referred to the clinic and evaluated for possible psychoeducational problems using a broad battery surveying domains of intelligence, memory, attention, academic achievement, and psychological/emotional functioning. All clinical participants consented to have their assessment information used confidentially in future research. This project was reviewed and approved by the Louisiana State University Institutional Review Board.

Individuals presented for assessment with a variety of primary complaints generally including depression or anxiety symptoms, learning problems, memory problems, and/or attentional concerns in the context of a university setting. As part of a comprehensive psychoeducational assessment, individuals were administered tests of intellectual functioning, memory functioning, attention/concentration, academic achievement, and psychological functioning. For the purposes of this study, only the testing data regarding intellectual and memory functioning were analyzed. Mean age in the clinical sample was 22.62 (*SD* = 6.80) with 513 (52.0%) males and 473 females (47.9%). Participant ethnic groups included Caucasian (*n* = 836, 84.7%), African American of (*n* = 112, 11.3%), Asian (*n* = 6, .6%), Hispanic (*n* = 18, 1.8%), Middle-Eastern (*n* = 6, .6%), and other (8, .8%). Demographic characteristics of the clinical sample regarding primary *DSM-IV-TR* diagnostic category are presented in Table 2. Additionally, 28.1% of those diagnosed with primary Axis I disorder also met criteria for another, co-occuring Axis I disorder (Table 3). Presentation of diagnostic rates in each clinical group (No External Incentive, External Incentive) is in Tables 4 and 5.

Procedures

For the experimental study, all measures were either administered by a senior level clinical

Table 2

Frequency and Percentage of Cases by Primary Diagnostic Category in the Clinical Group

	n	*%*
Depressive Disorders	75	7.6
Anxiety Disorders	114	11.6
Adjustment Disorders	12	1.2
Bipolar Disorders	15	1.5
Other Mood Disorders	4	.4
ADHD	236	23.9
Learning Disorders	173	17.5
Academic Problems (v-code)	12	1.2
Cognitive Disorders	84	8.5
Eating Disorders	2	0.2
Substance Use Disorders	9	0.9
Misc. V-Codes	5	0.5
Information Unavailable	5	0.5
Other	16	1.6
Diagnosis Deferred	15	1.5
No Diagnosis	209	21.2
Total	986	100

Table 3

Frequency and Percentage of Cases by Co-Occurring Secondary Diagnostic Category in the Clinical Group

	n	*%*
Depressive Disorders	48	4.7
Anxiety Disorders	87	8.8
Adjustment Disorders	10	1.0
Other Mood Disorders	4	0.4
ADHD	28	2.8
Learning Disorders	67	6.8
Academic Problems (v-code)	7	0.7
Cognitive Disorders	7	0.7
Eating Disorders	1	0.1
Substance Use Disorders	17	1.7
Misc. V-Codes	5	0.5
Other	2	0.2
Total	283	28.4

Table 4

Frequency and Percentage of Cases by Primary Diagnostic Category in the External Incentive Group

	n	*%*
Depressive Disorders	21	4.1
Anxiety Disorders	43	8.5
Adjustment Disorders	5	1
Bipolar Disorders	6	1.2
Other Mood Disorders	2	.4
ADHD	151	29.7
Learning Disorders	115	22.6
Academic Problems (v-code)	6	1.2
Cognitive Disorders	50	9.8
Eating Disorders	1	.2
Substance Use Disorders	2	.4
Misc. V-Codes	2	.2
Information Unavailable	2	.2
Other	9	1.8
Diagnosis Deferred	6	1.2
No Diagnosis	87	17.1
Total	508	100

Table 5

Frequency and Percentage of Cases by Primary Diagnostic Category in the No External Incentive Group

	n	*%*
Depressive Disorders	54	11.3
Anxiety Disorders	71	14.9
Adjustment Disorders	7	1.5
Bipolar Disorders	9	1.9
Other Mood Disorders	2	.4
ADHD	85	17.8
Learning Disorders	58	12.1
Academic Problems (v-code)	6	1.3
Cognitive Disorders	34	7.1
Eating Disorders	1	.2
Substance Use Disorders	7	1.5
Misc. V-Codes	3	.6
Information Unavailable	3	.6
Other	7	1.5
Diagnosis Deferred	9	1.9
No Diagnosis	122	25.5
Total	478	100

psychology graduate student specializing in neuropsychology or by an advanced undergraduate research assistant trained to administer the measures according to standardized instructions in the respective manuals. The research assistant was supervised on an ongoing basis to ensure maintained consistency and fidelity of test administration. The measures were given in one session, which took approximately two and a half hours to complete. While the administration order was counterbalanced during the experimental project, no such experimental control was possible for the clinical group. However, administration order effects for the WAIS-III and WMS-III are insignificant (Zhu & Tulsky, 2000), so strict counterbalancing was not a concern clinically.

For the clinical data, clinical psychology graduate students trained in test administration, theory, and application administered and scored all tests according to standardization procedures. As part of a full psychoeducational evaluation, clients were also administered a clinical interview. During this process, clients provided informed consent for psychological services, underwent full evaluation, and were provided with feedback regarding cognitive and emotional status that included diagnostic impressions.

Data Integrity

To ensure data integrity, completed protocols from the experimental project were carefully checked for errors and corrected along each of the three stages of data checking. That procedure consisted of 1) initially having trained undergraduate research assistants rescore and recalculate all subtest raw scores, 2) then converting all raw scores to standardized scores based on the national normative sample, and 3) finally entering those data into a computer spreadsheet. Through that process, it was discovered that there was at least one error in approximately 8.5% of the participants' data files, which were subsequently corrected. Correlations between the original dataset and the corrected dataset were all greater than .90 for every single variable in the

dataset, with an overall multiple correlation between original and corrected datasets exceeding .98. Data from the clinical and experimental samples were then checked for extreme values, outliers (four standard deviations away from the mean), data misentries, and any values not associated with particular tests. On the basis of that analysis, no participant was excluded.

Grouping Variables

In order to establish group status in the clinical sample, client charts were reviewed for their referral sources and reasons for referral. From that information, group determination was made according to each client's expressed reason for obtaining a psychoeducational evaluation. Accordingly, when the clinical demographic
information included clients' expressed interest in obtaining recommendations for academic accommodations and/or stimulant medication, they were categorized into the External Incentive group (n = 508). Of those in the External Incentive group, 428 (84.3%) were seeking academic accommodations, 50 (9.8%) were seeking medications, 30 (5.9%) were seeking medications and accommodations. If no such information was available to indicate such incentive seeking, the clients were designated into the No External Incentive group (n = 478). According to reliability coding analysis, 91% agreement of group status was obtained by independent raters. In the discrepant cases, the lead researcher (R.P.) independently reviewed the cases and determined group status to resolve conflicting values. Therefore, three groups were derived for this project: 1) Control, 2) No External Incentive, and 3) External Incentive.

Measures

Personality Assessment Inventory (PAI; Morey, 1991)

The PAI is a self-report objective measure of personality traits and clinical symptoms consisting of 344 statements on which participants can answer as *False*, *Slightly True*, *Mainly True*, or *Very True*. Each response form is computer scored and responses are compared with the

normative sample. Upon scoring, the PAI provides t-scores for four validity and nine clinical scales, with t-scores $\geq$ 70 indicating clinically notable elevations. The Negative Impression Management (NIM) validity scale is one of the most useful measures of symptom exaggeration (Boccaccini, Murrie, & Duncan, 2006; Calhoun, Earnst, Tucker, Kirby, & Beckham, 2000; Rogers, Sewell, Cruise, Wang, & Ustad, 1998). Rogers et al. (1998) suggested that a t-score of $\geq$ 77 correctly classifies 74% of honest responders and 84% of malingerers. Calhoun et al. (2000) reported using t-score cut ranging from 73 to 81. Boccaccini et al. (2006) indicated that using a cutoff of 81 in a forensic facility correctly identified 91% of malingerers who were diagnosed according to the Structured Inventory of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992). However, 70% of those misidentified as malingerers according to the NIM score performed in the "indeterminate" range of malingering on the SIRS, which is just below the cutoff for definite malingering and strongly suggests a negative response bias. Rogers, Sewell, Morey, and Ustad (1996) have also devised a discriminant function that has shown utility in differentiating malingered performance (Baity, Siefert, Chambers, & Blais, 2007; Sumanti, Boone, Savodnik, & Gorsuch, 2006). Moreover, the Malingering Index (Gaies, 1994) of the PAI has shown to be robust index of feigning in several samples (Blanchard, McGrath, Pogge, & Khadivi, 2003; Wang et al.).

Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997a)

Intellectual functioning was measured by the WAIS-III. In most cases, all subtests except the optional Object Assembly subtest were administered according to standardized protocol. Participants' raw scores were compared according to age with the national standardization sample and converted to scaled and standard scores according to standard manualized procedures as per the administration manuals (Wechsler, 1997a, 1997b)

Wechsler Memory Scale – Third Edition (WMS-III; Wechsler, 1997b)

Memory functioning was measured by the Wechsler Memory Scale-III. For the WMS-III, no optional subtests were administered. Since the WAIS-III and WMS-III contain identical Letter-Number Sequencing and Digit Span subtests, each subtest was administered only once on either measure.

Malingering Index Scores from the WAIS-III/WMS-III and PAI

Each index score for the measures was calculated according to methodology as described by the researcher noted below. In accordance with current recommendations (Bianchini et al., 2001; Greve & Bianchini, 2004), the author employed only those measures validated by at least one study using a known groups design. The following measures have also been shown to be useful in differentiating simulated noncredible performance as well (see above review). The following scores will be derived from the indices in Table 6 order to apply Slick et al.'s (1999) criteria to individual cases. Table 6 also contains cutoff scores used to categorize clinical patients according to the Slick Criteria for probable response bias. As recommended elsewhere (Babikian et al., 2006), evidence from two or more measures fulfilling Criterion B level evidence should suggest noncredible performance according to Slick Criteria. Therefore, performance on any two or more of the indices from the WAIS-III/WMS-III associated with noncredible performance are sufficient to meet Criterion B level evidence to support a finding of a *probable* noncredible case. Another method to classify as *probable* involves obtaining evidence from at least one of the WAIS-III/WMS-III indices *and* self-report evidence of a response bias from the PAI. As a result, several possible combinations are available to establish *probable* noncredible performance. To guard against false positive findings of noncredible performance, the author chose cutoff scores with high specificity (90% to 100%). In keeping with the conservative approach, any indices that

share subtest values in their calculation (e.g., Mittenberg Index and Reliable Digits) were only used to fulfill only one of the two B2 level findings per case.

Table 6

Cutoff Scores Corresponding to Index

Index	*Cutoff Score*
Reliable Digit Span (Greiffenstein et al., 1994)	≤ 6
Mittenberg Index (Mittenberg et al., 1995)	> .21
Vocabulary minus Digit Span (Mittenberg et al., 1995)	≥ 4
Maximum Digits Forward (Babikian et al., 2006; Heinly et al., 2005)	≤ 4
Age-Corrected Scale Score (Iverson, 1991; Iverson & Franzen, 1994)	≤ 4
Processing Speed Index (Etherton, Bianchini, Heinly et al., 2006)	≤ 70
Rarely Missed Index (Killgore & DellaPietra, 2000b)	≤ 40
WAIS-III Working Memory (Etherton, Bianchini, Ciota et al., 2006)	≤ 70
Faces I (Glassmire et al., 2003)	≤ 23
Auditory Delayed Recognition (Langeluddecke & Lucas, 2003)	≤ 42
Ord et al. Index (Ord et al., 2007)	≥ 3
Negative Impression Management t-score (Morey, 1991)	≥ 81t
Roger's Discriminant Function t-score (Rogers et al., 1996)	≥ 60t
Malingering Index (Morey, 1991)	≥ 3

Results

Hypothesis 1

Those in the External Incentive group will show significantly lower performance on the WAIS-III and WMS-III when compared to the No External Incentive and Control groups, particularly on measures of thinking speed and attention. WAIS-III FSIQ scores by group were in the average range: Control (M = 110.13, SD = 11.32), No External Incentive (M = 105.62, SD = 12.60), and External Incentive (M = 101.63, SD = 12.47). To test the first hypothesis, two separate one-way MANOVA's were conducted to determine if control and clinical groups differed on subtests from the measures of intelligence and memory. The first MANOVA included Group (Control, External Incentive, and No External Incentive) X WAIS-III subtests (Picture Completion, Vocabulary, Digit-Symbol Coding, Similarities, Block Design, Arithmetic, Picture Arrangement, Matrix Reasoning, Information, Digit Span, Letter-Number Sequencing, Comprehension, and Symbol Search). All power analyses were performed with G*Power 3.0.3 (Faul, Erdfelder, Lang, & Buchner, 2007). The number of participants needed for power, $(1 - \beta) = .95$, and significance level $\alpha = .05$ resulted in total sample size of 180 needed to detect a small effect (0.10). Upon a significant omnibus F test, follow-up univariate F-tests were conducted to determine which subtests differ between groups. A separate power analysis was conducted to ensure that the sample size would be adequate to detect a small/medium difference between the groups. Based on that analysis, a sample size of 690 would suffice.

The MANOVA revealed a significant between group main effect F (13, 1114) = 11.03, $p < .0001$, $\eta^2 = .116$. Results from that analysis with appropriate univariate and Bonferroni post-hoc follow-up procedures are presented in Table 7. As can be seen from the post-hoc comparisons, all predicted group differences fell in the expected direction with the External Incentive group scoring lower than the other two groups across all subtests.

Table 7

Univariate Analysis of Variance for WAIS-III Variables by Group (Clinical v. Control)

	Control (n = 182)	No External Incentive (n = 442)	External Incentive (n = 490)	F	p	Partial eta-squared
	Mean (SD)	*Mean (SD)*	*Mean (SD)*			
Vocabulary[abc]	13.01 (2.37)	11.77 (2.67)	11.10 (2.85)	33.51	.001	.057
Similarities[b]	11.61 (2.71)	11.03 (2.81)	10.64 (2.88)	8.09	.001	.014
Arithmetic[bc]	10.74 (2.36)	10.56 (2.60)	9.65 (2.93)	17.42	.001	.030
Digit Span[abc]	11.41 (2.73)	10.08 (2.78)	9.34 (2.70)	38.49	.001	.065
Information[bc]	11.70 (2.41)	11.39 (2.65)	10.60 (2.75)	15.83	.001	.028
Comprehension	11.91 (2.47)	11.66 (2.68)	11.37 (2.87)	3.01	.050	.005
Letter-Number Seq.[abc]	11.45 (2.72)	10.59 (2.62)	9.86 (2.64)	25.68	.001	.044
Picture Completion[bc]	10.73 (3.09)	10.54 (2.95)	10.06 (3.00)	4.72	.009	.008
Digit Symbol Cd[abc]	11.54 (2.53)	9.36 (2.51)	8.78 (2.65)	76.82	.001	.121
Block Design[bc]	11.20 (2.81)	10.79 (2.84)	10.24 (3.01)	8.64	.001	.015
Matrix Reasoning[bc]	12.21 (2.07)	11.91 (2.58)	11.34 (2.71)	9.91	.001	.018
Picture Arrangement	10.37 (2.75)	10.21 (2.64)	10.01 (2.82)	1.34	.263	.002
Symbol Search[abc]	12.04 (2.43)	10.06 (2.71)	9.39 (2.73)	64.88	.001	.105

[a] Control v. No External Incentive, [b]Control v. External Incentive, [c]External Incentive v. No External Incentive

Note. All Flagged Post-Hoc Comparisons, $p < .05$

However, based on those analyses, there were no significant group differences on the Comprehension or Picture Arrangement subtests. In keeping with the initial hypothesis, it was the measures of processing speed and attentional control that demonstrated the largest group differences (i.e., Digit Span, Letter-Number Sequencing, Digit-Symbol Coding, Symbol Search). Nonparametric analyses via the Wilcoxon Signed-Rank test also supported hypothesis one as lower scores were strongly associated with the External Incentive group as opposed to the No

External Incentive group, $W+$ (13) = 91, $p < .001$, and Control group, $W+$ (13) = 91, $p < .001$. Therefore, those clients explicitly seeking external gain demonstrated lower performance than controls and clients not reporting potential for external gain, particularly on the measures of processing speed and attentional control.

Similar analyses were conducted for the WMS-III subtests. Mean scores for the WMS-III General Memory Index by group were as follows: Control (M = 107.62, SD = 10.69), No External Incentive (M = 100.97, SD = 13.98), and External Incentive (M = 98.51, SD = 13.96). The second MANOVA included Group (Control, External Incentive, and No External Incentive) X WMS-III core subtests (Logical Memory I, Faces I, Verbal Paired Associates I, Family Pictures I, Spatial Span, Logical Memory II, Faces II, Verbal Paired Associates II, Family Pictures II). A power analysis was performed to determine the number of participants needed for power, $(1 - \beta) = .95$, and significance level, $\alpha = .05$ resulting in total sample size of 150 to detect a small effect (0.10). Upon a significant omnibus F, follow-up univariate F-tests were conducted to determine which subtests differ between groups. A separate power analysis was conducted to ensure that the sample size would be adequate to detect a small/medium difference (0.15) between the groups. Based on that analysis, a sample size of 690 was needed. The MANOVA revealed a significant between group main effect of group F (8, 1151) = 4.44, $p < .0001$, $\eta^2 = .03$.

Results from appropriate univariate and post-hoc follow-up procedures are presented in Table 8. While mean group performance on all of the measures were in the expected direction with the External Incentive group scoring lower than the other two groups, the External Incentive group only performed significantly lower than the No External Incentive and Control groups on the Logical Memory subtests, thus partially supporting the hypothesis. Nonparametric analyses via the Wilcoxon Signed-Rank test also supported the hypothesis as lower scores on the

Table 8

Univariate Analysis of Variance for WMS-III Variables by Group (Clinical v. Control)

	Control (n = 182)	No External Incentive (n = 466)	External Incentive (n = 503)	F	p	Partial eta-squared
	Mean (SD)	*Mean (SD)*	*Mean (SD)*			
Logical Memory I[abc]	11.13 (2.53)	10.09 (2.84)	9.55 (2.81)	21.81	.001	.037
Faces I[b]	10.38 (2.92)	9.93 (2.93)	9.60 (2.98)	4.88	.008	.008
Verbal Paired Ass. I[b]	10.67 (2.51)	10.32 (2.93)	9.98 (2.97)	4.18	.016	.007
Family Pictures I[ab]	10.90 (2.40)	10.05 (3.11)	9.82 (3.18)	8.43	.001	.014
Logical Memory II[abc]	11.76 (2.66)	10.41 (2.95)	9.84 (2.84)	30.19	.001	.050
Faces II[ab]	10.41 (2.58)	9.83 (2.67)	9.79 (2.79)	3.73	.024	.006
Verbal Paired Ass. II[ab]	11.19 (1.75)	10.59 (2.55)	10.42 (2.69)	6.33	.002	.011
Family Pictures II[ab]	10.81 (2.43)	9.85 (3.23)	9.64 (3.26)	9.54	.001	.016

[a] Control v. No External Incentive, [b]Control v. External Incentive, [c]External Incentive v. No External Incentive

Note. All Flagged Post-Hoc Comparisons, $p < .05$

WMS-III were strongly associated with the External Incentive group as opposed to the No External Incentive group, $W+ (8) = 36$, $p = .008$, and Control group, $W+ (8) = 36$, $p < .008$. In sum, there was strong support for the predicted group differences on intellectual functioning measures, but less parametric support for the same contrasts concerning the memory measures. Nevertheless, all scores fell in the predicted ordinal direction, a finding supported nonparametrically for both sets of measures.

Hypothesis 2

The External Incentive group will perform significantly different from the other two groups (No External Incentive and Control) on all validity indices, indicating a higher level of noncredible performance. To test the second hypothesis, an omnibus MANOVA was

conducted and followed by a series of one-way between subjects ANOVA's subsequent to a significant MANOVA finding. Group served as the independent variable and each of the examined WAIS-III/WMS-III validity indices were dependent variables. Given the results from the power analyses conducted above, the sample size obtained was adequate to detect a small to medium difference. The MANOVA revealed a significant between group main effect, F (11, 1100) = 12.34, $p < .0001$, $\eta^2 = .109$. Appropriate univariate and post-hoc follow-up procedures are presented in Table 9. Since the Ord et al. and Rarely Missed Indices from the WMS-III violated conservative conventions for acceptable skewness and/or kurtosis ($\leq$ -1 or $\geq$ 1), only nonparametric comparisons were conducted with those latter two variables.

Table 9

Univariate Analysis of Variance for WAIS-III/WMS-III Validity Indices by Group (Clinical v. Control)

	Control (n = 182)	No External Incentive (n = 435)	External Incentive (n = 487)	F	p	Partial eta-squared
	Mean (SD)	*Mean (SD)*	*Mean (SD)*			
WAIS-III WMI[abc]	106.79 (12.73)	102.25 (13.43)	97.38 (13.55)	36.48	.001	.062
Processing Speed Index[abc]	109.81 (12.30)	98.40 (12.82)	94.82 (12.99)	91.02	.001	.142
Auditory Recognition-Delayed Raw[abc]	50.81 (2.31)	49.80 (2.75)	49.30 (2.99)	19.67	.001	.035
WMS-III Faces I Raw[b]	39.23 (4.22)	38.23 (4.60)	37.57 (4.86)	8.63	.001	.015
Max. Digits Fwd.[abc]	7.20 (1.08)	6.72 (1.28)	6.50 (1.24)	21.16	.001	.037
Mittenberg Index	-.41 (.96)	-.40 (1.05)	-.25 (1.01)	3.03	.049	.005
Reliable Digits[abc]	10.95 (2.04)	10.04 (2.14)	9.40 (2.12)	37.39	.001	.064
Vocabulary-Digit Span	1.60 (2.76)	1.72 (3.33)	1.78 (3.36)	.20	.816	.001

[a] Control v. No External Incentive, [b]Control v. External Incentive, [c]External Incentive v. No External Incentive

Note. All Flagged Post-Hoc Comparisons $p < .05$

Omnibus one-way Kruskal-Wallis ANOVA revealed that group significantly affected the level of performance on the Ord et al. Index, χ^2 (2, 1152) = 33.27, $p < .0001$ (Table 10). In order to specify which groups differed according to the dependent variable, follow-up pairwise comparisons were conducted with Mann-Whitney *U*, controlling α for multiple comparisons, establishing significance at $p < .017$. In the first follow-up comparisons, control participants performed at a lower rate than those in No External Incentive clinical group, *U* (1, 648) = 36,415.00, $p < .0001$, and those in the External Incentive seeking clinical group, *U* (1, 685) = .37,069.00, $p < .0001$. However, the No External Incentive group and External Incentive seeking clinical groups did not differ, *U* (1, 969) = 111,992.50, $p = .07$. Medians, means, and standard deviations for the Ord et al. index are presented in Table 10. An omnibus one-way Kruskal-Wallis ANOVA revealed that group status was not significantly associated with the scores on the Rarely Missed Index, χ^2 (2, 1146) = .58, $p = .75$. Due to an insignificant initial omnibus finding, pairwise group comparisons were not conducted in order to protect against type I error. Medians, means, and standard deviations for the Rarely Missed Index by group are presented in Table 11.

Table 10

Median, Means, and Standard Deviations for the Ord et al. Index by Group (Clinical v. Control)

	Median	*Mean (SD)*
Control (n = 182)	.00	.05 (.28)
No External Incentive (n = 466)	.00	.30 (.75)
External Incentive (n = 503)	.00	.39 (.85)

Table 11

Median, Means, and Standard Deviations for the Rarely Missed Index by Group (Clinical v. Control)

	Median	*Mean (SD)*
Control (n = 182)	204.00	198.46 (24.62)
No External Incentive (n = 464)	204.00	195.99 (29.44)
External Incentive (n = 500)	204.00	194.44 (31.11)

There was general support for hypothesis three as the control participants scored significantly different than the No External Incentive and the External Incentive groups in the expected direction on five of ten validity indices. Further support for the hypothesis was also obtained via a Wilcoxon Signed-Rank analysis. Scores on the indices were associated with noncredible performance in the External Incentive group compared with the No External Incentive group, $W+ (10) = 55$, $p = .002$, and Control group, $W+ (10) = 55$, $p < .002$. Thus, there was a greater degree of noncredible performance associated with the External Incentive group rather than the other two groups. Additionally, the aggregated rank differences also strongly supported the dimensionality assertion in the following hypothesis, for which parametric statistics also provided moderate support.

Because indices from the PAI (Roger's Discriminant Function, Malingering Index, & Negative Impression Management) were only administered to the clinical participants, a separate MANOVA was conducted comparing only the External Incentive and no External Incentive groups. The MANOVA revealed a significant between group main effect, $F (3, 657) = 7.35$, $p < .0001$, $\eta^2 = .03$. While the External Incentive and No External Incentive group did not differ on the Malingering Index from the PAI, the External Incentive group scored significantly higher on

Roger's Discriminant function. Unexpectedly, the No External Incentive demonstrated significantly higher mean scores on the Negative Impression Management scale, counter to the author's hypothesis (Table 12). With the exception of the Negative Impression Management scale, all mean index scores again fell in the expected direction with the External Incentive group performing as predicted, supporting the overall hypothesis.

Table 12

Univariate Analysis of Variance for PAI Validity Indices by Clinical Group (No External Incentive v. External Incentive)

	No External Incentive (n = 305)	External Incentive (n = 356)
	Mean (SD)	*Mean (SD)*
Roger's Function**	50.15 (10.15)	53.20 (10.37)
Malingering Index	.79 (.87)	.78 (.89)
Neg. Impression Mgt.*	55.15 (12.12)	53.32 (11.03)

*$p < .05$., **$p < .001$

Hypothesis 3 & Classification of Noncredible Performance

The rate of noncredible performance as per Slick Criteria in the entire clinical sample is expected to be similar to rates in other general clinical settings where compensation seeking is a factor (20%) and the rate should be significantly higher among persons in the External Incentive group. While it was fairly common for control participants to score in the noncredible range on the Mittenberg Index and Vocabulary Minus Digit Span, failure on others was rare. Although means from the validity indices for the External Incentive clients reflected greater levels of noncredible performance than the control participants and No External Incentive group (see hypothesis two analyses), the proportion of clients falling in the noncredible range according

to published cutoff scores on each individual index differed only on three indices (Auditory Recognition-Delayed Raw, Mittenberg Index, & Reliable Digits) for the External Incentive group when compared with the No External Incentive group (Tables 13 & 14). This likely suggests that, while the External Incentive group performed to a greater degree in the direction of the noncredible range on each index, the difference may not be as clinically meaningful for some indices. Despite those observations from parametric findings, results from a Wilcoxon Signed-Rank test revealed that a higher proportion of those in the External Incentive group, compared to the No External Incentive group, scored in the noncredible range when examining test failure rate as a whole, $W+ (14) = 95$, $p < .005$. Thus the results reflected the predicted ordinality of the scores by group.

In order to further assess group differences according to failure rates on validity indices, the author also examined the proportion of clients failing at least one measure. 46.3% of the External Incentive group failed at least one index from the WAIS-III/WMS-III, indicating *possible* noncredible performance according to the Slick Criteria (i.e., only one level of Criterion B evidence), demonstrating that it was fairly common to fail one or more index. In contrast, 36.8% of the No External Incentive group failed one or more, which is less than the External Incentive group (46.3%), $\chi^2 (1) = 9.0$, $p = .003$. Table 15 contains the cumulative percentage of WAIS-III/WMS-III validity indices failed per group. Next, the author presents data regarding multiple failures. Despite the high failure rates on the Mittenberg Index and Vocabulary Minus Digit Span, along with some failures on other indices in the control group, none of the control participants' combination of index scores met Criterion B of the Slick Criteria, resulting in 100% specificity for the current classification scheme (i.e., no false positives). The rate of noncredible performance according to Slick Criteria, when accounting for all indices (including the PAI), was lower than expected ($n = 59$, 12.1% of the External Incentive group, 5.98% of the entire

Table 13

Participants Scoring beyond Cutoff Scores for Validity Indices by Group*

	Control	No External Incentive	External Incentive
	n (%)	*n (%)*	*n (%)*
WAIS-III WMI	0 (0)	2 (0.4)	6 (1.2)
Processing Speed Index	0 (0)	4 (0.9)	6 (1.2)
Digit Span Scale Score	0 (0)	4 (0.9)	9 (1.8)
Auditory Recognition-Delayed Raw[b]	0 (0)	3 (0.6)	15 (3.0)
WMS-III Faces I Raw	0 (0)	1 (0.2)	5 (1.0)
Max. Digits Fwd.	3 (1.6)	10 (2.2)	16 (3.2)
Mittenberg Index[b]	42 (23.1)	125 (28)	170 (34.1)
Reliable Digits[b]	2 (1.1)	12 (2.6)	32 (6.4)
Vocabulary-Digit Span	40 (22)	131 (27.3)	148 (29.5)
Ord et al. Index	0 (0)	12 (2.6)	20 (4.0)
Rarely Missed Index	0 (0)	2 (0.4)	1 (0.2)
Roger's Function[a]	---	59 (19.3)	81 (22.8)
Malingering Index[a]	---	12 (3.9)	16 (4.5)
Neg. Impression Mgt[a]	---	20 (6.6)	16 (4.5)

*Reflects percentage of participants in analyses.
[a]Index not administered to control participants.
[b]All Flagged Chi-Square Comparisons: No External Incentive vs. External Incentive, $p < .05$.

Table 14

Participants Scoring beyond Cutoff Scores for Validity Indices for Control and Total Clinical Group*

	Control	Clinical
	n (%)	*n (%)*
WAIS-III WMI	0 (0)	8 (0.8)
Processing Speed Index	0 (0)	10 (1.0)
Digit Span Scale Score	0 (0)	13 (1.4)
Auditory Recognition-Delayed Raw	0 (0)	18 (1.9)
WMS-III Faces I Raw	0 (0)	6 (0.6)
Max. Digits Fwd.	3 (1.6)	26 (2.7)
Mittenberg Index	42 (23.1)	295 (31.2)
Reliable Digits	2 (1.1)	44 (4.6)
Vocabulary-Digit Span	40 (22)	279 (29.1)
Ord et al. Index	0 (0)	32 (3.3)
Rarely Missed Index	0 (0)	3 (0.3)
Roger's Function[a]	---	140 (21.2)
Malingering Index[a]	---	28 (4.2)
Neg. Impression Mgt[a]	---	36 (5.4)

*Reflects percentage of participants in analyses..
[a]Index not administered to control participants.

Table 15

Number of WAIS-III/WMS-III Indices Failed according to Percentage of Participants beyond Cutoff Scores

Number of Indices Failed	Control	No External Incentive	External Incentive	Total Clinical
	n (%)	*n (%)*	*n (%)*	*n (%)*
0	123 (67.6)	302 (63.2)	273 (53.7)	575 (58.3)
1	35 (19.2)	75 (15.7)	102 (20.1)	177 (18.0)
2	21 (11.5)	79 (16.5)	96 (18.9)	175 (17.7)
3	2 (1.1)	13 (2.7)	17 (3.3)	30 (3.0)
4	1 (0.5)	7 (1.5)	15 (3.0)	22 (2.2)
5	---	2 (0.4)	2 (0.4)	4 (0.4)
6		---	2 (0.4)	2 (0.2)
7			1 (0.2)	1 (0.1)
8			---	---

clinical sample). Please recall that those in the No External Incentive group did not satisfy Criterion A (presence of a substantial external incentive) of the Slick Criteria, thus precluding them from meeting full Slick Criteria for MND. However, members of the No External Incentive group (n = 42, 9.7%) did have index combinations that satisfied Slick Criterion B (see classification scheme in the Methods section), which was not significantly different from the External Incentive group (12.1%), χ^2 (1) = 2.1, p = .14. However, this lack of finding is likely due to the imperfect classification methodology of group assignment. When using the WAIS-III/WMS-III indices alone, the rate of positive Criterion B findings in the External Incentive group dropped to just 17 (3.5%) and 9 (0.9%) in the No External Incentive group, χ^2 (1) = 2.1,

p = .15. Therefore, just 3% (n = 26) of the entire clinical group met Criterion B evidence when only using the WAIS-III/WMS-III. As a result, excluding the PAI indices in the present analyses would have resulted in not identifying at least a 72.2% of those meeting Slick Criteria in this sample. Regardless of incentive level, 10.2% of the entire clinical sample met Criterion B of the Slick Criteria when combining failure rates across the WAIS-III/WMS-III and PAI indices (see Table 16 for description of Criterion B rate according to group), which was lower than expected.

Table 16

Frequency of Participants by Group Satisfying Criterion B from the Slick Criteria according to Index Type

	WAIS-III/WMS-III Indices Only	All Indices
	n (%)	*n (%)*
Control*	0 (0%)	N/A
No External Incentive	9 (0.9%)	42 (9.7)
External Incentive	17 (3.5)	59 (12.1)

*Control Participants not Administered PAI

Concordance of clinical behavioral observations and Slick Criteria follow. Of the 59 clients in the probable MND group, behavioral observations made by the evaluating clinician during the assessment indicated "fidgety" behavior (n = 1), giving up easily (n = 1), variable effort (n = 3), anxiety during the session (n = 2), frustration during the session (n = 1), and fatigue at the time of evaluation (n = 2); no other descriptors of irregular behavior were made. Thus, in at least 48 (81%) of the probable noncredible cases, clients were not behaviorally identified as putting forth questionably valid performance even though the combination of embedded validity measures indicated otherwise, highlighting importance of including measures of validity in psychoeducational evaluations and replicating and amplifying the cautions

espoused by Faust and Ziskin (1988) regarding the failure of clinicians to be able to identify noncredible performance based on performance observations alone.

Hypothesis 4

It is hypothesized that those with self-reported pre-existing cognitive disorders will have a lower rate of noncredible performance as they have previously established that they have cognitive deficits in order to obtain external gain. Therefore, they may not feel the need to exaggerate their deficits. To test the assumption that self-reported history of diagnosed learning problems, special education, cognitive disorders, or diagnosed attentional conditions would not be associated with increased failure on the indices, Chi Square Tests of Independence were conducted with history of self-reported cognitive problems as the independent measure and proportion of failure on each of the indices as dependent measures.

As can be seen in Table 17, such self-report was not associated with noncredible performance on individual indices after controlling for multiple comparisons by adjusting the significance level ($p < .0035$). Nevertheless, results from a Wilcoxon Signed-Rank test revealed that, overall, failure on validity indices was actually associated with those reporting a history of cognitive problems compared to those not reporting such a history, $W+ (14) = 83$, $p < .05$. Also counter to hypothesis four, this variable was associated with performing in the probable noncredible range according to Slick Criteria as 25 (4.6%) of those without a history of cognitive problems fell into the probable noncredible range while 34 (7.7%) of those with such self-reported problems met criteria for probable noncredible performance, $\chi^2 (1) = 4.14$, $p = .042$. Therefore, those with a history of cognitive problems tended to fail the indices and their pattern and rates of failure also resulted in higher incidents of Criterion B classification. As a result, this hypothesis was not generally supported, but there were conflicting findings.

Table 17

Rates of Noncredible Performance by History of Cognitive Problems*

Index	History of Cognitive Problems (n = 440)	No History of Cognitive Problems (n = 541)
	n (%)	*n (%)*
Working Memory ≤ 70	5 (1.2)	3 (0.6)
Processing Speed ≤ 70	5 (1.6)	5 (1.0)
Digit Span Scaled Score ≤ 4[a]	10 (2.3)	3 (0.6)
Aud. Rec. Raw ≤ 42	11 (2.6)	7 (1.3)
Faces I Raw ≤ 23	2 (0.5)	4 (0.8)
Max. Digits Fwd. ≤ 4	15 (3.5)	11 (2.1)
Mittenberg Index > .21[a]	150 (35.1)	145 (28.2)
Reliable Digits ≤ 6	26 (6.1)	18 (3.4)
Voc.-Digit Span ≥ 4	128 (29.7)	150 (28.7)
Ord et al. Index ≥ 3[a]	20 (4.6)	12 (2.3)
Rarely Missed Index ≤ 40	0 (0.0)	3 (0.6)
Roger's Function ≥ 60t[a]	76 (24.9)	63 (17.8)
Malingering Index ≥ 3	11 (3.6)	17 (4.8)
Neg. Imp. Mgt ≥ 81t[a]	10 (3.3)	26 (7.4)

*Reflects percentage of participants in analyses.
[a]$p < .05$

Hypothesis 5

It is expected that those meeting Slick Criteria for probable MND in the current study will be diagnosed with current brain based disorders (e.g., Cognitive Disorder NOS, ADHD, LD) most frequently as an outcome of their current psychoeducational evaluation. This hypothesis was investigated by subdividing the probable MND group and remaining clinical group according to *DSM-IV-TR* diagnostic category. As can be seen in Table 18, 79.7% (n = 47) of those meeting Slick Criteria were diagnosed with some mental disorder according to their psychoeducational evaluation, which was slightly higher, but not significantly different from that observed in the remainder of the total clinical sample (n = 693, 74.8%), χ^2 (1) = .71, p = .40. However, the percentage of the remaining clinical group diagnosed with a psychiatric diagnosis was higher (n = 239, 25.8%) than those in the MND group (n = 8, 13.6%), χ^2 (1) = 4.4, p = .036. In contrast, the rate of being diagnosed with a brain based disorder in the probable MND group (n = 38, 64.4%) *was* higher than the remainder of the entire clinical group (49.1%, n = 455), χ^2 (1) = 5.2, p = .02. Thus, while the rate of diagnoses did not appear to be higher in the probable MND group compared with the remainder of the clinical group as a whole, the rate of disorders specifically related to cognitive dysfunction was higher. Support of this hypothesis suggests that those meeting Slick Criteria in this sample obtained a disproportionate rate of cognitive diagnoses despite strong evidence from the embedded measures that they likely exaggerated deficits. Information regarding diagnostic comorbidity in the MND group is presented in Table 19.

Related to Hypothesis Five is the observance that the frequency of psychiatric diagnoses in the No External Incentive (n = 158, 33%) group was higher than the External Incentive group (n = 89, 17.5%), χ^2 (1) = 31.7, p < .001. However, in mirroring the findings of the MND analyses, the proportion of those diagnosed with a brain-based disorder in the External

Table 18

Frequency and Percentage of Primary Axis I Diagnoses in the Probable MND Group

	n	%
Major Depressive Episode	1	1.7
Generalized Anxiety Disorder	1	1.7
Bipolar Disorder	1	1.7
Obsessive Compulsive Disorder	1	1.7
Cyclothymic Disorder	1	1.7
Adjustment Disorder	1	1.7
Academic Problem	1	1.7
Autism Spectrum Disorder	1	1.7
Information Unavailable	1	1.7
Anxiety Disorder, NOS	2	3.4
Reading Disorder	2	3.4
Math Disorder	2	3.4
Disorder of Written Expression	2	3.4
Diagnosis Deferred	2	3.4
Amnestic Disorder	4	6.8
Learning Disorder, NOS	7	11.9
Cognitive Disorder, NOS	8	13.6
No Diagnosis	8	13.6
ADHD	13	22

Table 19

Frequency and Percentage of Cases by Co-Occurring Secondary Diagnostic Category in the Probable MND Group

	n	%
Depressive Disorders	2	3.4
Anxiety Disorders	6	10.2
Adjustment Disorders	1	1.7
Schizoaffective Disorder	1	1.7
ADHD	4	6.8
Learning Disorders	3	5.1
Misc. V-Codes	2	3.4

Incentive (n = 316, 62.2%) group was higher than the No External Incentive group (n = 183, 38.3%), χ^2 (1) = 56.4, p < .001. Taken together, the results suggest that even overtly seeking an external incentive in a psychoeducational evaluation is associated with obtaining a diagnosis for a brain-based disorder.

Hypothesis 6

This hypothesis predicts that those clients meeting Slick Criteria for probable MND were likely to be recommended to receive academic accommodations and/or be recommended for a medication consultation at a higher rate than those not meeting criteria for probable MND. The author tested hypothesis six via Chi Square Test of Independence by calculating the number of individuals recommended to receive academic accommodations and/or medication. The author first compared proportions of recommendations between the External Incentive (n = 343, 71.0%) and No External Incentive group (n = 211, 47.4%), which indicated a significant finding, χ^2 (1) = 53.61, p = .00001. The next analysis compared the derived probable MND group

and remaining combined clinical groups according to receiving recommendations for medication referral and/or academic accommodations by using Chi Square Test of Independence. According to that analysis, 43 (75.4%) of those in the probable group and 511 (58.7%) of the remaining clinical patients received one of those recommendations, χ^2 (1) = 6.25, p = .012. However, those meeting Slick Criteria for MND did not receive those recommendations at a higher rate than those in the remaining External Incentive group (n = 300, 70%), suggesting the recommendation of accommodations might more likely be related to participants' status as members of the External Incentive group than their actual classification as probable MND.

Hypothesis 7

The specificity levels of each validity index will be greater when those meeting the Slick Criteria are compared to the Control group rather than when compared to the No External Incentive group and remaining External Incentive group. Sensitivity and specificity levels were calculated according to proportion of those falling above and below cutoff scores per group based on comparisons between the probable MND and the control group (Table 20), No External Incentive group (Table 21), and the remaining External Incentive group (non-probable MND clinical group; Table 22). Values for positive predictive power are provided for each index according to several theoretical base rate conditions (10%, 20%, 30%, 40%, 50%) by using the calculations outlined by Baldessarini et al. (1983).

The greatest margin of specificity differences between the groups was with Reliable Digits, which is also the most well validated measure of validity among the indices and the most specific to detecting noncredible performance. That measure was also followed by the Working Memory Index and Processing Speed Index indicators as the most specific to probable MND in this sample. As can be seen from the tables, specificity levels for each index were higher in the Control group than both clinical groups, supporting the hypothesis. Specificity was also lower, as

a whole, in the No External Incentive group when compared with the External Incentive group after removing probable MND clients, W- (13) = 65.5, p = .02. This latter finding was surprising due to the fact that external incentive was shown, in the previous analyses, to be related to testing scores and validity index scores. An increased failure rate was expected due to the fact that the remainder of the External Incentive group presumably had motivation to feign deficits. This finding may reflect an artifact such that experimental group assignment may not have achieved pure groups or may be related to other factors noted in the discussion. Other observations from the tables indicate that when obtaining a failed index score, the probability for noncredible performance is increased, especially in conditions where the base rate exceeds 20%. However, the diagnostic confidence in any one failed measure decreases as a direct function of decreasing hypothetical base rates as has been shown elsewhere (Gouvier, 1999).

Table 20

Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for Control Group

			Positive Predictive Power for Base Rates				
	Spec	Sens	.10	.20	.30	.40	.50
Working Memory ≤ 70	100	6.8	100	100	100	100	100
Processing Speed ≤ 70	100	15.3	100	100	100	100	100
Digit Span Scaled Score ≤ 4	100	8.5	100	100	100	100	100
Aud. Rec. Raw ≤ 42	100	11.9	100	100	100	100	100
Faces I Raw ≤ 23	100	1.7	100	100	100	100	100
Max. Digits Fwd. ≤ 4	98.4	6.8	32	52	65	74	81
Mittenberg Index > .21	76.9	86.4	29	48	62	71	79
Reliable Digits ≤ 6	98.1	79.7	82	91	95	97	98
Voc.-Digit Span ≥ 4	88	39	27	45	58	68	76
Ord et al. Index ≥ 3	100	15.2	100	100	100	100	100
Rarely Missed Index ≤ 40	100	0	100	100	100	100	100

Table 21

Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for No External Incentive

			Positive Predictive Power for Base Rates				
	Spec	Sens	.10	.20	.30	.40	.50
Working Memory ≤ 70	99.6	6.8	65	81	88	92	94
Processing Speed ≤ 70	99.1	15.3	65	81	88	92	94
Digit Span Scaled Score ≤ 4	99.1	8.5	51	70	80	86	90
Aud. Rec. Raw ≤ 42	99.4	11.9	69	83	89	93	95
Faces I Raw ≤ 23	98.2	1.7	9	19	29	39	49
Max. Digits Fwd. ≤ 4	97.8	6.8	26	44	57	67	76
Mittenberg Index > .21	72	86.4	26	44	57	67	76
Reliable Digits ≤ 6	97.4	79.7	77	88	93	95	97
Voc.-Digit Span ≥ 4	72.7	39	14	26	38	49	59
Ord et al. Index ≥ 3	97.4	15.2	39	59	71	80	85
Rarely Missed Index ≤ 40	99.4	0	0	0	0	0	0
Roger's Function ≥ 60t[a]	80.7	65.5	27	46	59	69	77
Malingering Index ≥ 3[a]	96.1	14.5	29	48	61	71	79
Neg. Imp. Mgt ≥ 81t[a]	93.4	20	25	43	56	67	75

[a]Index not administered to control participants.

Table 22

Sensitivity, Specificity, and Positive Predictive Power for Validity Indices for External Incentive - Not Probable Group

			Positive Predictive Power for Base Rates				
	Spec	Sens	.10	.20	.30	.40	.50
Working Memory ≤ 70	99.5	6.8	60	77	85	90	93
Processing Speed ≤ 70	100	15.3	100	100	100	100	100
Digit Span Scaled Score ≤ 4	99.1	8.5	51	70	80	86	90
Aud. Rec. Raw ≤ 42	98.2	11.9	42	62	74	82	87
Faces I Raw ≤ 23	99.1	1.7	17	32	45	56	65
Max. Digits Fwd. ≤ 4	97.3	6.8	22	39	52	63	72
Mittenberg Index > .21	73	86.4	26	44	58	68	76
Reliable Digits ≤ 6	95.5	79.7	66	82	88	92	95
Voc.-Digit Span ≥ 4	73.8	39	14	27	39	50	60
Ord et al. Index ≥ 3	97.5	15.2	40	60	72	80	86
Rarely Missed Index ≤ 40	99.9	0	0	0	0	0	0
Roger's Function ≥ 60t[a]	85	65.5	33	52	65	74	81
Malingering Index ≥ 3[a]	97.3	14.5	37	57	70	78	84
Neg. Imp. Mgt ≥ 81t[a]	98.3	20	57	75	83	89	92

[a]Index not administered to control participants.

Discussion

Performance validity is crucial to consider when conducting and interpreting neuropsychological testing (Bianchini et al., 2001; Slick et al., 1999). The assessment of client credibility has been especially stressed when results can be used to demonstrate evidence for disability status and/or to receive external gain (Bush et al., 2005). Traditionally, feigned symptoms or deficit/symptom exaggeration has been a concern in the context of litigation, workers' compensation, and other situations that may result in financial rewards or relief from occupational or other duties (Mittenberg et al., 2002). More recently, focus has been turned to investigate the context of potentially exaggerated neurocognitive deficits in other situations as well (Delis & Wetter, 2007). In the current study, the author investigated the relationship of overtly seeking academic accommodations and/or medication referrals and neurocognitive performance in a university setting through several hypotheses.

Hypothesis 1

Those in the External Incentive group will show significantly lower performance on the WAIS-III and WMS-III when compared to the No External Incentive and Control groups, particularly on measures of thinking speed and attention. This hypothesis was generally supported as the External Incentive group performed significantly lower than the No External Incentive group on 10 of 13 subtests from the WAIS-III. Although the External Incentive group did score lower than the No External Incentive group on the remaining three subtests (Similarities, Comprehension, & Picture Arrangement), the difference did not reach statistical significance using parametric tests of significance. Comparison of the External Incentive group with the Control group also showed significantly lower performance on 11 of 13 WAIS-III subtests with the exception of Comprehension and Picture Arrangement. All comparisons fell in the expected direction nonparametrically, further supporting this hypothesis.

In contrast to the WAIS-III comparisons, the External Incentive group performed significantly lower than the No External Incentive group on only the Logical Memory subtests when contrasting WMS-III memory performance. While the External Incentive group performed significantly lower on only the Logical Memory subtests in that comparison, they did show lower absolute scores across all WMS-III subtests, albeit the differences did not rise to statistical significance parametrically. An additional follow-up Wilcoxon signed rank test did indicate that, on the whole, the External Incentive group scores on the WMS-III were significantly consistently lower than those with no explicit motivation for external secondary gain and controls.

The parametric finding that the External Incentive group performed significantly different from the other clinical group on only Logical Memory in the WMS-III comparison was not particularly surprising due to the nature of the referrals to the clinic and the face validity of the measure. For example, the overwhelming majority of referrals to the clinic were sought for attentional complaints and/or academic difficulties. Since the WMS-III appears as an obvious memory measure, clients seeking to exaggerate neurocognitive deficits within the present sample did not likely view the WMS-III as a measure relevant to their alleged current deficits, thus deficits would not be exaggerated on those tests. This is supported by the fact that the largest group differences among the WAIS-III subtests related to attentional functioning (i.e., Digit Span, Letter-Number Sequencing, Digit-Symbol Coding, and Symbol Search).

In contrast, the Logical Memory subtests in particular, requires a high level of attentional skill and cognitive load and may be more likely viewed by the client as a measure relevant to attentional assessment. Therefore, the two memory measures on the WMS-III that appear most similar to attentional functioning were performed at the lowest level among clients explicitly seeking evaluation as a means to obtain an external incentive. Despite this, all scores fell in the hypothesized direction which resulted in significant nonparametric findings when considering

score ranks between groups across all measures of intelligence and memory.

Hypothesis 2

Hypothesis two proposed that the External Incentive group will perform significantly different from the other two groups (No External Incentive and Control) on all validity indices, indicating a higher level of noncredible performance. Hypothesis two was imperfectly but strongly supported with the External Incentive group performing significantly different from the No External Incentive group on half of the 10 indices from the WAIS-III/WMS-III. All remaining five parametrically nonsignificant findings with this comparison fell in the expected direction and the hypothesis was fully supported through nonparametric analyses. Similarly, the External Incentive group performed significantly different on seven of the 10 indices when contrasted with the Control group, with all comparisons falling in the expected direction, indicating a higher degree of noncredible performance. Again, all nonsignificant findings also fell in the expected direction with additional follow-up Wilcoxon signed rank tests indicating that the External Incentive group scores on the WMS-III validity indices were lower than the No External Incentive group. Additionally, the No Incentive group also performed significantly toward the invalid range on six of 10 WAIS-III/WMS-III validity indices when compared to the Control group. The results add validity evidence for the embedded measures as the controls performed in the direction of invalidity to a significantly lesser extent than clients with and without external incentive. Furthermore, those with stated incentive for external gain performed more often in the noncredible range to a significantly greater degree than those not seeking some form of expressively stated secondary gain.

As for the PAI validity indices, the External Incentive group performed significantly higher than the No External Incentive group only on the Roger's Discriminant Function. Contrary to expectations, the No External Incentive group mean scores on the Negative

Impression Management was higher than the External Incentive group, indicating a higher level of noncredible performance on that particular index. There was no group difference in the Malingering Index from the PAI. Thus, there was less support for the hypothesis utilizing the PAI indices, especially in a standalone capacity.

Overall, none of the mean scores from any one validity index across all groups fell in the noncredible range, indicating the relatively infrequent occurrence of the extreme degree of performance required for classification as noncredible. Despite one contradictory finding, the majority of results supported the general assertion that those overtly seeking external incentives scored differently than both Controls and those clients not overtly seeking external incentives. In sum, with additional support from the first hypothesis, these results suggest that clients' explicit motivation for testing influences neurocognitive test performance and failure rates on validity indices as well. Not only that, but it also appears that those without explicit external incentive seeking behaviors in this context score differently on validity indices than those from the normal population, which will be addressed below.

Hypothesis 3 & Classification of Noncredible Performance

The rate of noncredible performance as per Slick Criteria in the entire combined clinical sample is expected to be similar to the rates in other general clinical settings where compensation seeking is a factor (20%) and the rate should be significantly higher among persons in the External Incentive group. It was generally expected that the proportion of noncredible findings on validity indices would be similar to a recent observation in a similar sample (Sullivan et al., 2007). Overall, 46% of the clients in the External Incentive group met Slick Criteria for at least possible MND because they failed at least one index from the WAIS-III/WMS-III. This was greater than in the No External Incentive group (36.8%), which was on par with recent reports of failure rates of four embedded validity measures in neuropsychological

patients not meeting Slick Criteria (Victor, Boone, Serpa, Buehler, & Zeigler, In Press). A surprise finding among the control group was a comparably high rate of validity test failures on individual measures (32.4%), particularly on the Mittenberg Index and Vocabulary Minus Digit Span which runs counter to what was expected, especially considering this study employed conservative cutoff scores (Inman & Berry, 2002; Orey et al., 2000). In all, the failure rates across groups in the current study were above the single index failure rate (22%) from the Word Memory Test as reported by Sullivan et al. (2007). The overall profile of failure rates supported the nonparametric dimensionality as hypothesized with the highest failure in the External Incentive group followed by the No External Incentive and then Control participants.

Despite the seemingly high failure rate on individual indices, Hypothesis Three was not fully supported as the proportion of clients in the entire clinical group satisfying Slick Criteria for probable MND was only 6%. This figure did rise to 12% as a proportion of the External Incentive group alone, though it was far below the expected 50% noted in other settings with a high rate of external incentive seeking (Mittenberg et al., 2002). Therefore, although groups with an increased prior probability for malingering contain a higher number of participants meeting Slick Criteria, the number here did not rise to traditionally high levels consonant with other highly saturated groups (i.e., defense referrals). Beyond the issue of meeting full criteria for probable MND (seeking external incentive), 9.7% of the No External Incentive group also satisfied criterion B of the Slick Criteria (as outlined in the Methods section) when the PAI was included in the decision matrix. As a whole then, 10.2% of the entire clinical group, irrespective of incentive seeking status, met criterion B of the Slick Criteria. At minimum, this strongly suggests the regular identification of some form of noncredible performance on neurocognitive testing even in those not seeking overtly stated external incentives. However, that does not necessarily imply malingered performance *per se*. On the other hand, despite the fact that 32.4%

off the controls failed at least one validity index, no control participant met Criterion B of the Slick Criteria. Consequently, no control participant was falsely identified as a malingerer.

The rate of suspected MND in this sample decreased dramatically when using only the WAIS-III/WMS-III validity indices as the means with which to meet Slick Criteria (3.5% of the External Incentive group, 0.9% of the No External Incentive group). Therefore, only 3% of the total sample met Criterion B without using the PAI indices, suggesting a high false negative rate when relying solely on the Wechsler scales to detect noncredible performance.

Due to the fact that the rates of probable MND were lower than expected, and lower than other current estimates in similar samples (Sullivan et al., 2007), it is assumed to be an underestimate. However, Sullivan et al. (2007) did not specifically address the issue of MND as per Slick Criteria, but simply reported failure rates on singular indices from the Word Memory Test. Therefore, their report of 22% failure rates do not reflect derivation of malingered behavior. Though failure rates on multiple indices from the Word Memory Test were likely substantially lower in that study, they were not reported. Moreover, they did not control for external incentive seeking and their level may indicate a 22% rate of possible MND, but it is unclear because they did not account for compensation seeking.

One reason for the somewhat low rate of probable MND in the current sample is the classification scheme employed in this study utilized a very stringent and conservative methodology to categorize clients according to the probable MND Slick Criteria (Boone, 2007b; Larrabee, Greiffenstein, Greve, & Bianchini, 2007). This included using indices with above 90% specificity as empirically established in brain injured groups, samples with significant cognitive impairment beyond what is typically associated with specific LD and ADHD and certainly more impairment than university students (Belanger, Curtiss, Demery, Lebowitz, & Vanderploeg, 2005; Belanger & Vanderploeg, 2005). Hence, failure rates in the current sample are not likely

due to the mild level of impairment typically served in university clinics.

The author also used multiple indicators according to varying principles (e.g., self-report, floor effect, performance curve methodologies) to determine noncredible performance that guarded against capitalizing on chance for meeting Criterion B of the Slick Criteria by requiring failure on two or more nonoverlapping validity indices (Rosenfeld, Sands, & Van Gorp, 2000). Furthermore, the lower rate of probable MND in this sample occurred despite the fact that some of the current measures (Mittenberg Index and Vocabulary-Digit Span) performed with unacceptably low specificity in the comparable control sample when contrasted with the extant literature, even when using conservative cutoff scores. As a result, using either the Mittenberg Index or Vocabulary-Digit Span in isolation in samples similar to those used in the present study is not recommended due to decreased specificity. Indeed, using only the WAIS-III/WMS-III, 32.4% of the Control group scored in the noncredible range on at least one validity indicator, representing low specificity (high false positive rate), which highlights the need to use multiple validity indicators (Larrabee, 2007a; Meyers & Volbrecht, 2003). However, singular indices may serve utility as initial, preliminary screening measures of invalid test performance due to the literature base that supports their usage. In this way, noting failure on one measure increases the post-test odds of noncredibility, but does not confirm it. Nonetheless, the reason for high failure rates on individual indices in the Control sample and No External Incentive group remains somewhat unclear in this study especially considering patients with marked neurological patients were excluded from this study. One explanation is that a proportion of clients classified in the No External Incentive group actually had an incentive, but did not overtly make that known at the time of the evaluation. Thus, the substantially lowered proportion meeting Criterion B likely represents the lower limit of noncredible performance in this population, and those meeting probable criteria for MND represent the extreme cases of MND.

Hypothesis 4

It is hypothesized that those with self-reported pre-existing cognitive disorders will have a lower rate of noncredible performance as they have previously established that they have cognitive deficits in order to obtain external gain. Therefore, they may not feel the need to exaggerate their deficits. The author also investigated the relationship of a self-reported history of diagnosed cognitive problems and scoring below the cutoff on validity indices. It was expected that those with such a history would not score below cutoff criteria on the validity measures at a higher rate than those without that history. However, those with a history of cognitive problems were more likely to present for testing to obtain accommodations at the university and/or for standardized "high stakes" placement tests (e.g., American College Test, etc.). Therefore, the external incentive to perform poorly on the neurocognitive testing was quite salient in this subgroup of clients. As such, those with a history of cognitive problems and scholastic struggles were actually more likely to meet Slick Criteria for probable MND than those without that background. This may be partly due to previous involvement in neurocognitive testing, resulting in familiarity with the fact that one needs to demonstrate (or exaggerate) deficits in order to receive a diagnosis. As a result of experience with testing situations and previous external gains as a function of that testing and disorder familiarity, the effect may be similar to being coached as to the types of deficits needed to be accentuated to obtain a diagnosis. It is thought that this possibility may be a function of time since diagnoses. For example, the majority of individuals in the pre-existing diagnosis group may have recently obtained diagnoses rather than being diagnosed in childhood. It was rationally derived that those diagnosed in childhood would expect professions to retain their cognitive diagnosis by virtue of long-standing problems and they would not feel the need to exaggerate deficits. Therefore, in the future, it would be important to analyze differences in validity failure rates as a function of time

since diagnoses (i.e., diagnosis rendered in childhood versus adulthood). To conclude, those clients with pre-existing diagnoses presenting to clinics to establish a diagnosis to receive external secondary gain may have a higher prior probability of exaggerating or feigning deficits, contrary to this hypothesis, and raises the question of whether many of these previously diagnosed individuals obtained their original diagnoses via simulation.

Hypothesis 5

It is expected that those meeting Slick Criteria for probable MND will be diagnosed with current brain based disorders (e.g., Cognitive Disorder NOS, ADHD, LD) most frequently as an outcome of their current psychoeducational evaluation. Initially, it was not thought that this hypothesis was supported because the rate of any primary Axis I mental disorders in the probable MND group (79.7%) was very similar to the rate in those not meeting Slick Criteria (75%). However, a more fine-grained, theory-driven reanalysis showed that the rate of being diagnosed with a brain-based disorder (e.g., ADHD, LD, Cognitive Disorder) according to their current psychoeducational evaluation was increased in the probable MND group (64%) compared with those not meeting probable MND criteria (49%). Therefore, those meeting criteria for probable MND obtained a brain-based diagnosis more often than those demonstrating wholly credible performance. In particular, those meeting Slick Criteria were successful in feigning/exaggerating deficits to the extent that they obtained diagnoses to justify receiving external gain.

Hypothesis 6

Those clients meeting criteria for probable MND were likely to be recommended to receive academic accommodations and/or be recommended for a stimulant medication consultation at a higher rate than those not meeting criteria for probable MND. Related to hypothesis five, there was an association for clinical group status as 71% of the External

Incentive group and 47% of those in the No External Incentive clinical group received such recommendations based on current psychoeducational results. Moreover, those meeting Slick Criteria for probable MND had a 75.4% rate of receiving recommendations as opposed to 58.7% of the clients not meeting Slick Criteria.

As with the first, second, third, and fourth hypotheses, there was support for an association of clinical group status (No External Incentive & External Incentive) with outcomes from current testing and diagnoses. In this case, it was shown that the group of clients visiting the clinic with explicit intentions of obtaining secondary gain did in fact receive a high rate of desired recommendations. This finding was higher than expected given the low rates of probable MND, suggesting that the estimation of probable MND was actually too low (high number of false negatives). The sixth hypothesis also supported the assertion that those meeting probable Slick Criteria actually succeeded in demonstrating a neurocognitive impairment profile that warranted formal recommendations for academic accommodations and/or stimulant medication referral. Furthermore, 70% of those in the External Incentive group (with the MND clients removed) obtained desired recommendations, indicating a strong effect for seeking external gain. Only a small percentage of those patients were identified as putting forth questionable effort at the time of the evaluation via behavioral observations and none were diagnosed with MND. In sum, university students seeking incentives appear to be able exaggerate symptoms to obtain desired outcomes, and they are not likely identified without symptom validity testing.

Hypothesis 7

The specificity levels would be greater when those meeting the Slick Criteria for probable MND are compared to the Control group rather than when compared to the No External Incentive group and remaining External Incentive group. Following the findings that the No External Incentive group failed the validity indices at a higher rate than the Control

sample, specificity levels for each index followed this pattern (see Tables 19, 20, & 21). Despite the fact that the instruments were validated with brain injured samples to establish cutoff criteria, research has also shown higher failure rates in credible clinical samples compared with nonpatient samples (Victor et al., In Press), which was also demonstrated by differential specificity values in the No External Incentive and the clients seeking external incentive, but not meeting Slick Criteria for MND. As a result of the varying specificity levels, predictive values of the tests were altered with increasing classification certainty obtained when utilizing the Control group as the comparison sample (see Table 20). The two poorest predictive values in the Control participant comparison were those with the highest failure rate (Mittenberg Index and Vocabulary minus Digits) as well.

Those two indices share validation samples in original research that did not include an adequate proportion of participants with normal and high intellectual quotients as were present in the current university sample. Since both of these indices are essentially based on utilizing a subtest discrepancy analysis, the high failure rates in the current study may reflect problems with applying that approach in high functioning samples. For instance, it has been widely noted that subtest discrepancies on intelligence scales are more common in those with above average FSIQs (Hawkins & Tulsky, 2001; Saklofske, Tulsky, Wilkins, & Weiss, 2003). While the Vocabulary minus Digit Span index has been investigated in a large normative sample (Iverson & Tulsky, 2003), failure rates have not been stratified by overall intellectual level, which likely results in underestimating the false positive rate of that index. No such normative comparisons with the Mittenberg Index exist for WAIS-III standardization sample, and attempts to obtain the relevant data have been unsuccessful (Psychological Corporation, personal communication, November 2007). Therefore, other data regarding the appropriate applicability of the two discrepancy based validity indices to high FSIQ samples are lacking. Despite the high failure on those measures in

normal controls, all other indices demonstrated excellent specificity ranging from 98.1 to 1.00.

In the current sample, there are at least three other reasons for why the specificity and predictive values differed when comparing the MND group with the No External Incentive group and the remaining credible clients in the External Incentive group and those rather than the Control group. First, it may be that there was an actual increased "false positive" rate in the No External Incentive group and the portion of those with external incentive that did not meet full Criterion B evidence. Interpretation from other research illuminates this interpretation if the operational definition of a false positive is declaring someone a malingerer based on Slick Criteria. In this case, the test objectively over-identifies those who actually demonstrate credible performance in the context of seeking an external gain as well as those without such motives. Second, because of the retrospective group assignment, the "false positive" cases may actually be true positives that were misassigned to the No External Incentive group, precluding the application of the Slick Criteria to those clients (not meeting Criterion A). Thus, those clients classified by testing as "false positives" clients in the No external Incentive group are actually false negative cases due to the quasi-experimental manipulation and not due to the psychometric characteristics of the failed validity indices, thus being initial classification errors. For example, if a client was seeking external incentive, but did not explicitly state that, he or she would have been misclassified into the No External Incentive group in the current study, contaminating the purity of the experimental groups. Conversely, the established psychometric properties of the indices and previous research may actually over identify credible patients, which would indicate that the false positive findings in the No External Incentive group identified true cases of MND.

The last possibility, and perhaps the most overlooked in the research on MND, is that the No External Incentive group contained several individuals seeking evaluation based on *a priori* set of psychological needs to demonstrate personal pathology. Thus, the high "false positive"

rates reflect a tendency to present with somatoform and/or factitious features, rendering the testing noncredible. While Delis and Wetter (2007) and others (Binder, 2007; Boone, 2007a; Larrabee, 2007b) have recognized that factors other than substantial external secondary gain may relate to feigned neurocognitive performance, the area remains understudied. This is particularly the case regarding the explanation of so-called false positive rates of validity tests in patient samples with no apparent external incentives from participating in the evaluation process. In this manner, false positive results are viewed as test properties while excluding other more recent conceptualizations of somatic and cognitive symptom feigning, resulting in "overshadowing" alternative explanations. Thus, the false positive rates reported in patient samples may reflect an underlying cogniform syndrome that drives failures on neuropsychological validity tests and not simply psychometric noise. This is potentially problematic given the high rates of suspected noncredibility across settings, as effort and financial incentives have been shown to account for significant variance in neuropsychological test performance (Binder & Rohling, 1996; Green et al., 2001) and there are few guidelines in place to aid the clinician when deciding how to interpret neurocognitive performance in the presence of failed validity testing *and* no apparent external incentive.

Limitations

Although the general findings indicated that the External Incentive group performed more poorly than the No External Incentive group on the subtests from the WAIS-III/WMS-III, not all comparisons were statistically significant when applying parametric techniques. However, nonparametrically, the significant overall tendency was that the External Incentive group performed lower than the No External Incentive group much more often. The External Incentive group scored significantly differently than the No External Incentive group on the validity indices in an imperfectly consistent manner, though nonparametric results indicate the overall

tendency to score in the direction noncredible direction.

Thus, the first two hypotheses were supported and some of the mild inconsistencies may have been an experimental artifact in the grouping variables. For instance, the level of external incentive seeking was determined via retrospective chart reviews wherein information regarding the clients' true reason for obtaining the evaluation may not have been, at times, explicitly stated. Therefore, in those cases, the author grouped those clients into the No External Incentive group as a means to avoid mislabeling those individuals as meeting criteria for MND in keeping with the conservative approach in this study. Please note that in order to meet Slick Criteria, one has to have external incentive (Criterion A). Because of this, the No External Incentive group likely contained individuals who actually had an external incentive (i.e., seeking accommodations or medications), but did not directly indicate their motivation at the time of evaluation, resulting in a contaminated group. As a result, participants with external incentive were likely classified as not having external incentive. Moreover, those misclassified clients could not meet Slick Criteria as a function of that group assignment. However, that experimental artifact may actually bolster the robustness of the significant findings because it is thought that those individuals that may have been misclassified into the No External Incentive group should have performed similarly to clients in the External Incentive group thereby diluting the group differences, which likely only happened to a small degree. This potential situation would also be similar to typical clinical practice where the clinician is often uncertain as to the clients' explicit motivation for evaluation. Additionally, it is also possible that a client's explicit denial of external incentives may not be credible because they may refuse to admit such motivational factors.

Despite the significant group differences, it may also be possible that those in the External Incentive group may have performed lower on the WAIS-III/WMS-III due to differences related to actual functioning. For instance, it could be that those with *bonafide*

neurocognitive problems more readily expressed their intentions to receive external gain, and were categorized as such. Be that as it may, the External Incentive group still performed toward the noncredible range on the validity indices more often than the No External Incentive group, which calls into question that possibility.

Another limitation of this study relates to the use of validity measures validated for use in a medico-/psycholegal context and with neurological groups that have largely not included clients diagnosed with ADHD or LD. Therefore, the appropriateness of assigning labels of noncredible performance or malingering to clients in the present sample may be premature. To counteract those concerns, the author utilized conservative cutoff scores to minimize false positive findings of noncredible performance based on specificity levels across a number of empirical findings.

The high number of control participants failing the Mittenberg Index and Vocabulary minus Digit Span does call into question the use of cutscores for those measures in a university sample. Each of those two indices is based on score discrepancies within the intellectual domain, which may partially explain their low specificity in this sample given the increasing occurrence of such discrepancies as IQ scores become higher and higher. For instance, it has been noted that individuals with high intellectual quotients are more likely to show less uniform intellectual abilities than those with low or average functioning (Hawkins & Tulsky, 2001). As a result, large and frequent score discrepancies across intellectual abilities (i.e., a high degree of subtest scatter) are expected in those with above average intelligence. Given that the control participants were comprised of well-educated university students, the tendency for positive findings on the Mittenberg Index and Vocabulary minus Digit Span may have reflected true discrepancies in scores rather than noncredible performance. However, it has been shown that more extreme performance on those validity indices indicates a greater degree of test invalidity (Greve et al.,

2003; Mittenberg et al., 2001), which was also supported by the group comparisons but also suggests that the cutoff score in this population may need modification for future use. This topic also highlights that reliance on the observation of any single validity indicator is likely to result in false findings of MND (Boone, 2007b; Bush et al., 2005; Larrabee, 2007a; Larrabee et al., 2007; Meyers & Volbrecht, 2003; Rosenfeld et al., 2000; Victor et al., In Press). Therefore, the use of a multiple methods and validity tests should be employed to guard against undue stigmatization.

Perhaps the greatest limitation of the current study is that there was no independently validated patient group meeting Slick Criteria using measures outside the scope of this study. Because of this, it is not possible to definitively state that those meeting Slick Criteria according to the current methodology were, in fact, feigning deficits in order to receive an external gain. Nonetheless, the number of individuals meeting Slick Criteria in this study did not appear excessive and was actually lower than expected given estimates of noncredible performance in other clinical settings using samples motivated by different external incentives For instance, according to Mittenberg et al. (2002), the rate of MND in patients not involved with litigation or seeking compensation approximates 7% and the percentage in samples with high rates of external incentives approaches 50%. The percentage of MND found in the current study falls well below the high rate and just above the low estimate for MND in other clinical samples, suggesting that the rate in the current study were reasonable.

While the current study did not include an independently validated group of patients meeting research criteria for MND, the current literature base has not provided much technical guidance in this area for the current population of interest. As noted above, nearly all of the research on noncredible neurocognitive performance is concerned with forensic cases having identifiable high stakes outcomes that affect public policy, service provision, and public as well

as private financial considerations. In fact, mention of nonforensic malingering in particular has been sparse and vague (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998) making it unclear if current models of malingering and noncredible performance in forensic and other settings are applicable to nonforensic clinical evaluations with less sensational external reasons for seeking assessment. Therefore, there are no readily available instruments or methods that directly apply to determine credible performance specifically for those seeking ADHD or LD evaluations, which reflects the preliminary nature of this study and this area of research.

Conclusions

Neuropsychologists spend the majority of their time outside the realm of forensic practice (Kanauss, Schatz, & Puente, 2005) and several sources of secondary gain independent of the legal arena remain unstudied. The results from this project support extending research and clinical investigation of noncredible performance to the academic setting as it was shown that there was a relationship between overtly seeking non-financial external incentives and test performance. More specifically, as a group, university students who overtly seek academic accommodations and/or stimulant medications performed lower on neuropsychological measures and showed more failure rates on validity indices than those not overtly seeking such incentives.

As a result, incentive-seeking can be conceptualized on a continuum that pertains to influence of external incentives that includes the perception, desirability, and type of external incentive. For instance, failure rates on neuropsychological validity indices and performance on neuropsychological tests can be examined based on the degree of incentive present (financial, employment status, non-financial awards, etc.). As it relates to the current study, further investigations would be well-served to examine the various types of external academic incentives (i.e., medication seeking, accommodation seeking) and neuropsychological performance.

So too, external pressures experienced by clients also need to be examined for an effect

and interaction with incentive levels. For instance, in a forensic context, an independently wealthy claimant may not view a potential legal $1,000 award as enticing as someone who was in debt and laid off of work. In terms of academics, students with high pressure to perform scholastically (i.e., low GPA, students on academic probation) may view particular academic accommodations as being differentially attractive versus honor students with high GPA. Therefore, it will be necessary to further refine and more clearly operationalize how *substantial* particular external incentives may be viewed in the context of neuropsychological evaluations. In this vein, a theoretically-driven approach to studying noncredible performance in the context of external incentives may be offered in order to elucidate otherwise difficult to explain poor neuropsychological performance.

References

Advokat, C. D., Guidry, D., & Martino, L. (2008). Licit and illicit use of medications for Attention-Deficit Hyperactivity Disorder in undergraduate college students. *Journal of American College Health, 56*(6), 601-606.

APA. (2000). *Diagnostic and statistical manual of mental disorders (4th ed.-Text Revision)*. Washington, DC: Author.

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*(1), 84-94.

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist, 21*(6), 899-916.

Ashendorf, L., O'Bryant, S. E., & McCaffrey, R. J. (2003). Specificity of malingering detection strategies in older adults using the CVLT and WCST. *Clinical Neuropsychologist, 17*(2), 255.

Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with Digit Span from the Wechsler Adult Intelligence Scale-Third Edition. *The Clinical Neuropsychologist, 20*(3), 513-523.

Axelrod, B. N., & Rawlings, D. B. (1999). Clinical utility of incomplete effort WAIS-R formulas: A longitudinal examination of individuals with traumatic brain injury. *Journal of Forensic Neuropsychology, 1*(2), 15-27.

Babcock, Q., & Byme, T. (2000). Student perceptions of methylphenidate abuse at a public liberal arts college. *Journal of American College Health, 49*, 143 -145.

Babikian, T., & Boone, K. B. (2007). Intelligence tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 103-127). New York, NY: The Guilford Press.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *Clinical Neuropsychologist, 20*(1), 145-159.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning & Verbal Behavior, 9*(2), 176-189.

Baity, M. R., Siefert, C. J., Chambers, A., & Blais, M. A. (2007). Deceptiveness on the PAI: A study of naïve faking with psychiatric inpatients. *Journal of Personality Assessment, 88*(1), 16-24.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*(5), 569-573.

Bar-Hillel, M., & Hogarth, R. M. (1990). Back to base rates. In *Insights in decision making: A tribute to Hillel J. Einhorn.* (pp. 200-216). Chicago, IL, US: University of Chicago Press.

Belanger, H. G., Curtiss, G., Demery, J. A., Lebowitz, B. K., & Vanderploeg, R. D. (2005). Factors moderating neuropsychological outcomes following mild traumatic brain injury: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(3), 215-227.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*(4), 345-357.

Benton, A. L., Sivan, A. B., Hamsher, K., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment (2nd Ed). A clinical manual*. New York, NY: Oxford University Press.

Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12*(5), 715-728.

Bernard, L. C., McGrath, M. J., & Houston, W. (1993). Discriminating between simulated malingering and closed head injury on the Wechsler memory scale-revised. *Archives of Clinical Neuropsychology, 8*(6), 539-551.

Bianchini, K. J., Love, J. M., Greve, K. W., & Adams, D. (2005). Detection and diagnosis of malingering in electrical injury. *Archives of Clinical Neuropsychology, 20*(3), 365-373.

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

Bigler, E. D. (2006). Can author bias be determined in forensic neuropsychology research published in Archives of Clinical Neuropsychology? *Archives of Clinical Neuropsychology, 21*(5), 503-508.

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

Binder, L. M. (2007). Comment on Cogniform Disorder and Cogniform Condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22*(6), 681-682.

Binder, L. M., & Rohling, M. L. (1996). Money matters: a meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry, 153*(1), 7-10.

Binder, L. M., Villanueva, M. R., Howieson, D., & Moore, R. T. (1993). The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. *Archives of Clinical Neuropsychology, 8*(2), 137-147.

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3*(2), 175-181.

Blanchard, D. D., McGrath, R. E., Pogge, D. L., & Khadivi, A. (2003). A comparison of the PAI and MMPI-2 as predictors of faking bad in college students. *Journal of Personality Assessment, 80*(2), 197-205.

Boccaccini, M. T., Murrie, D. C., & Duncan, S. A. (2006). Screening for malingering in a criminal-forensic sample with the Personality Assessment Inventory. *Psychological Assessment, 18*(4), 415-423.

Boone, K. (2007a). Commentary on "Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms" by Dean C. Delis and Spencer R. Wetter. *Archives of Clinical Neuropsychology, 22*(6), 675-679.

Boone, K. (2007b). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 29-49). New York, NY: The Guilford Press.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20*(4), 419-426.

Butters, N., Salmon, D. P., Cullum, C. M., & Cairns, P. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale--Revised. *The Clinical Neuropsychologist, 2*(2), 133-148.

Calhoun, P. S., Earnst, K. S., Tucker, D. D., Kirby, A. C., & Beckham, J. C. (2000). Feigning combat-related Posttraumatic Stress Disorder on the Personality Assessment Inventory. *Journal of Personality Assessment, 75*(2), 338-350.

Carlesimo, G. A., Fadda, L., Lorusso, S., & Caltagirone, C. (1994). Verbal and spatial memory spans in Alzheimer's and multi-infarct dementia. *Acta Neurologica Scandinavica, 89*(2), 132-138.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability Consultative Exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*(1), 1-14.

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*(2), 191-198.

Conti, R. P. (2004). Malingered ADHD in adolescents diagnosed with conduct disorder: A brief note. *Psychological Reports, 94*(3), 987-988.

Crank, J. N., & Deshler, D. D. (2001). Disability eligibility issues and university student assessment outcomes. *Journal of Vocational Rehabilitation, 16*, 217-226.

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2007). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist*, 1-18.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (DKEFS): Examiner's manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., & Wetter, S. R. (2007). Cogniform Disorder and Cogniform Condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22*(5), 589-604.

Demakis, G. J. (2004). Application of clinically-derived malingering cutoffs on the California Verbal Learning Test and the Wechsler Adult Intelligence Test-Revised to an analog malingering study. *Applied Neuropsychology, 11*(4), 220-226.

Duncan, S. A., & Ausborn, D. L. (2002). The use of reliable digits to detect malingering in a criminal forensic pretrial population. *Assessment, 9*(1), 56-61.

Etherton, J. L., Bianchini, K. J., Ciota, M. A., & Greve, K. W. (2005). Reliable digit span is unaffected by laboratory-induced pain: Implications for clinical use. *Assessment, 12*(1), 101-106.

Etherton, J. L., Bianchini, K. J., Ciota, M. A., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering and the WAIS-III Working Memory Index. *The Spine Journal, 6*(1), 61-71.

Etherton, J. L., Bianchini, K. J., Greve, K. W., & Heinly, M. T. (2005). Sensitivity and Specificity of Reliable Digit Span in malingered pain-related disability. *Assessment, 12*(2), 130-136.

Etherton, J. L., Bianchini, K. J., Heinly, M. T., & Greve, K. W. (2006). Pain, malingering, and performance on the WAIS-III Processing Speed Index. *Journal of Clinical & Experimental Neuropsychology, 28*(7), 1218-1237.

Evans, S. W., Serpell, Z. N., Schultz, B. K., & Pastor, D. A. (2007). Cumulative benefits of secondary school-based treatment of students with Attention Deficit Hyperactivity Disorder. *School Psychology Review, 36*(2), 256-273.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Faust, D., Guilmette, T. J., Hart, K., Arkes, H. R., Fishburne, F. J., & Davey, L. (1988). Neuropsychologists' training, experience, and judgment accuracy. *Archives of Clinical Neuropsychology, 3*(2), 145-163.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: the capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*(4), 578-582.

Faust, D., & Nurcombe, B. (1989). Improving the accuracy of clinical judgment. *Psychiatry, 52*(2), 197-208.

Faust, D., & Ziskin, J. (1988). The expert witness in psychological and psychiatry. *Science, 241*, 31-35.

Fisher, D. C., Ledbetter, M. F., Cohen, N. J., Marmor, D., & Tulsky, D. S. (2000). WAIS-III and WMS-III profiles of mildly to severely brain-injured patients. *Applied Neuropsychology, 7*(3), 126-132.

Ford, C. V. (1983). *The somatizing disorders: illness as a way of life*. New York, NY: Elsevier.

Frederick, R. I. (1997). *Validity Indicator Profile manual*. Minnetonka, MN: NCS Assessments.

Frederick, R. I., Sarfaty, S. D., Johnston, J. D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology, 8*(1), 118-125.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Gaies, L. A. (1994). *Malingering of depression on the Personality Assessment Inventory.* Unpublished doctoral dissertation, University of South Florida, Tampa, US.

Glassmire, D. M., Bierley, R. A., Wisniewski, A. M., Greene, R. L., Kennedy, J. E., & Date, E. (2003). Using the WMS-III faces subtest to detect malingered memory impairment. *Journal of Clinical & Experimental Neuropsychology, 25*(4), 465.

Gouvier, W. D. (1999). Base rates and clinical decision making in neuropsychology. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice.* (pp. 27-37). Lisse, Netherlands: Swets & Zeitlinger Publishers.

Gouvier, W. D. (2001). Are you sure you're really telling the truth? *NeuroRehabilitation, 16*(4), 215-219.

Gouvier, W. D., Cubic, B., Jones, G., & Brantley, P. (1992). Postconcussion symptoms and daily stress in normal and head-injured college populations. *Archives of Clinical Neuropsychology, 7*(3), 193-211.

Gouvier, W. D., Hayes, J. S., & Smiroldo, B. B. (1998). The significance of base rates, test sensitivity, test specificity, and subjects' knowledge of symptoms in assessing TBI sequelae and malingering. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation.* (pp. 55-79). New York, NY, US: Plenum Press.

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

Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., et al. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: preliminary findings. *The Clinical Neuropsychologist, 21*(6), 929-942.

Green, P. (2003). *Green's Word Memory Test for Microsoft Windows*. Edmonton, Alberta: Green's Publishing Inc.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 40-77). New York, NY: The Guilford Press.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M., III. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*(12), 1045-1060.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*(3), 218-224.

Greve, K. W., & Bianchini, K. J. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: a methodological commentary with recommendations. *Archives of Clinical Neuropsychology, 19*(4), 533-541.

Greve, K. W., Bianchini, K. J., Mathias, C. W., Houston, R. J., & Crouch, J. A. (2003). Detecting malingered performance on the Wechsler Adult Intelligence Scale: Validation of Mittenberg's approach in traumatic brain injury. *Archives of Clinical Neuropsychology, 18*(3), 245-260.

Greve, K. W., Springer, S., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., et al. (2007). Malingering in toxic exposure: Classification accuracy of reliable digit span and WAIS-III digit span scaled scores. *Assessment, 14*(1), 12-21.

Hadley, W. M. (2007). The necessity of academic accommodations for first-year college students with learning disabilities. *Journal of College Admission,* 195, 9-13.

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

Hawkins, K. A., & Tulsky, D. S. (2001). The influence of IQ stratification on WAIS-III/WMS-III FSIQ-General Memory Index discrepancy base-rates in the standardization sample. *Journal of the International Neuropsychological Society, 7*(7), 875-880.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*(5), 892-900.

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS digit span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment, 12*(4), 429-444.

Henry, G. (2005). Probable malingering and performance on the Test of Variables of Attention. *The Clinical Neuropsychologist, 19*(1), 121-129.

Hilsabeck, R. C., LeCompte, D. C., Marks, A. R., & Grafman, J. (2001). The Word Completion Memory Test (WCMT): A new test to detect malingered memory deficits. *Archives of Clinical Neuropsychology, 16*, 669-677.

Hilsabeck, R. C., Thompson, M. D., Irby, J. W., Adams, R. L., Scott, J. G., & Gouvier, W. D. (2003). Partial cross-validation of the Wechsler Memory Scale-Revised (WMS-R) General Memory-Attention/Concentration Malingering Index in a nonlitigating sample. *Archives of Clinical Neuropsychology, 18*(1), 71-79.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 967-974.

Inman, T. H., & Berry, D. T. R. (2002). Cross-validation of indicators of malingering: a comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology, 17*(1), 1-23.

Iverson, G. L. (1991). *Detecting malingered memory deficits through the use of multiple objective measures: A preliminary investigation*: Unpublished master's thesis, West Virginia University, Morgantown.

Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 15*(2), 829-858.

Iverson, G. L., & Franzen, M. D. (1994). The Recognition Memory Test, digit span, and Knox Cube Test as markers of malingered memory impairment. *Assessment, 1*(4), 323-334.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18*(1), 38-51.

Iverson, G. L., Slick, D. J., & Franzen, M. D. (2000). Evaluation of a WMS-R malingering index in a non-litigating clinical sample. *Journal of Clinical & Experimental Neuropsychology, 22*(2), 191.

Iverson, G. L., & Tulsky, D. S. (2003). Detecting malingering on the WAIS-III: Unusual Digit Span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology, 18*(1), 1-9.

Johnson, S. K., Lange, G., DeLuca, J., Korn, L. R., & Natelson, B. (1997). The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis, and depression. *Applied Neuropsychology, 4*(3), 145-153.

Johnstone, L., & Cooke, D. J. (2003). Feigned intellectual deficits on the Wechsler Adult Intelligence Scale-Revised. *British Journal of Clinical Psychology, 42*(3), 303-318.

Kanauss, K., Schatz, P., & Puente, A. E. (2005). Current trends in the reimbursement of professional neuropsychological services. *Archives of Clinical Neuropsychology, 20*(3), 341-353.

Killgore, W. D. S., & DellaPietra, L. (2000a). Item response biases on the logical memory delayed recognition subtest of the Wechsler Memory Scale-III. *Psychological Reports, 86*(3 Pt 1), 851-857.

Killgore, W. D. S., & DellaPietra, L. (2000b). Using the WMS-III to detect malingering: Empirical validation of the Rarely Missed Index (RMI). *Journal of Clinical & Experimental Neuropsychology, 22*(6), 761-771.

Kirmayer, L. J., & Sartorius, N. (2007). Cultural models and somatic syndromes. *Psychosomatic Medicine, 69*(9), 832-840.

Kumho Tire v. Carmichael, 525 U.S. 137, 152 (1999).

Lange, R. T., Iverson, G. L., Sullivan, K., & Anderson, D. (2006). Suppressed Working Memory on the WMS-III as a marker for poor effort. *Journal of Clinical & Experimental Neuropsychology, 28*(3), 294-305.

Langeluddecke, P. M., & Lucas, S. K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale--Third edition in mild head injury litigants. *Archives of Clinical Neuropsychology, 18*(2), 181-197.

Langeluddecke, P. M., & Lucas, S. K. (2004). Validation of the Rarely Missed Index (RMI) in detecting memory malingering in mild head injury litigants. *Journal of Forensic Neuropsychology, 4*(1), 49-64.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *Clinical Neuropsychologist, 12*(2), 179-188.

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

Larrabee, G. J. (2003b). Exaggerated pain report in litigants with Malingered Neurocognitive Dysfunction. *Clinical Neuropsychologist, 17*(3), 395-401.

Larrabee, G. J. (2007a). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 1 - 14.

Larrabee, G. J. (2007b). Commentary on Delis and Wetter, "Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms". *Archives of Clinical Neuropsychology, 22*(6), 683-687.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits.* (pp. 334-371). New York, NY, US: Oxford University Press.

Leark, R. A., Dixon, D., Hoffman, T., & Huynh, D. (2002). Fake bad test response bias effects on the test of variables of attention. *Archives of Clinical Neuropsychology, 17*(4), 335-342.

Lees-Haley, P. R., Earnest, K., & Dolezal-Wood, S. (1995). Post-Concussive symptoms: Base rates and etiology in psychiatric patients. *The Clinical Neuropsychologist, 9*(1), 89-92.

Lees-Haley, P. R., Smith, H. H., Williams, C. W., & Dunn, J. T. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology, 11*(1), 45-51.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment (4th ed.)*. New York, NY, US: Oxford University Press.

Lu, P. H., Boone, K. B., Jimenez, N., & Razani, J. (2004). Failure to inhibit the reading response on the Stroop Test: a pathognomonic indicator of suspect effort. *Journal of Clinical & Experimental Neuropsychology, 26*(2), 180-189.

Marshall, P., & Happe, M. (2007). The performance of individuals with mental retardation on cognitive tests assessing effort and motivation. *The Clinical Neuropsychologist, 21*(5), 826 - 840.

Mathias, C. W., Greve, K. W., Bianchini, K. J., Houston, R. J., & Crouch, J. A. (2002). Detecting malingered neurocognitive dysfunction using the Reliable Digit Span in traumatic brain injury. *Assessment, 9*(3), 301-308.

McCabe, S. E., Teter, C. J., & Boyd, C. J. (2006). Medical use, illicit use, and diversion of prescription stimulant medication. *Journal of Psychoactive Drugs, 38*(1), 43-56.

McCabe, S. E., Teter, C. J., Boyd, C. J., & Guthrie, S. K. (2004). Prevalence and correlates of illicit methylphenidate use among 8th, 10th, and 12th grade students in the United States, 2001. *Journal of Adolescent Health, 35*(6), 501-504.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30*(5), 525-564.

Meyers, J. E., & Volbrecht, M. (1998). Validation of reliable digits for detection of malingering. *Assessment, 5*(3), 303-307.

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

Miller, L. J., Ryan, J. J., Carruthers, C. A., & Cluff, R. B. (2004). Brief screening indexes for malingering: A confirmation of vocabulary minus digit span from the WAIS-III and the rarely missed index from the WMS-III. *Clinical Neuropsychologist, 18*(2), 327-333.

Millis, S. R. (2004). Evaluation of malingered neurocognitive disorders. In M. R. P. J. Esslinger (Ed.), *Principles and practice of behavioral neurology and neuropsychology* (pp. 1077–1089). Philadelphia, PA: W.B. Saunders Company.

Millis, S. R., Ross, S. R., & Ricker, J. H. (1998). Detection of incomplete effort on the Wechsler Adult Intelligence Scale-Revised: a cross-validation. *Journal of Clinical and Experimental Neuropsychology, 20*(2), 167-173.

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment (2nd ed.)*. New York, NY, US: Oxford University Press.

Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale--Revised. *Psychological Assessment, 5*(1), 34-40.

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

Mittenberg, W., Theroux-Fichera, S., Zielinski, R., & Heilbronner, R. L. (1995). Identification of malingered head injury on the Wechsler Adult Intelligence Scale--Revised. *Professional Psychology: Research and Practice, 26*(5), 491-498.

Mittenberg, W., Theroux, S., Aguila-Puentes, G., Bianchini, K. J., Greve, K. W., & Rayls, K. (2001). Identification of malingered head injury on the Wechsler Adult IntelligenceScale-3rd Edition. *Clinical Neuropsychologist, 15*(4), 440.

Moline, S., & Frankenberger, W. (2001). Use of stimulant medication for treatment of Attention-Deficit/Hyperactivity Disorder: A survey of middle and high school students' attitudes. *Psychology in the Schools, 38*, 569-584.

Morey, L. C. (1991). *Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources, Inc.

Murphy, J. M. (1976). Psychiatric labeling in cross-cultural perspective. *Science, 191*(4231), 1019-1028.

Nichols, E., Harrison, A. G., McCloskey, L., & Weintraub, L. (2002). *Learning opportunities task force 1997-2002: Final report*: Richmond Hill: Learning Opportunities Task Force.

Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9*(6), 501-552.

Ord, J. S., Greve, K. W., & Bianchini, K. J. (2007). Using the Wechsler Memory Scale-III to detect malingering in mild traumatic brain injury. *The Clinical Neuropsychologist*, 1-16.

Orey, S. A., Cragar, D. E., & Berry, D. T. R. (2000). The effects of two motivational manipulations on the neuropsychological performance of mildly head-injured college students. *Archives of Clinical Neuropsychology, 15*(4), 335-348.

Osmon, D., Plambeck, E., Klein, L., & Mano, Q. (2006). The Word Reading Test of Effort in adult learning disability: A simulation study. *The Clinical Neuropsychologist, 20*(2), 315-324.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology, 47*(2), 409-410.

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

Parker, T. (2000). Timmy 2000. In F.C. Agnone, P. Bady, A. Garefino, J. Howell, D. Leibling, T. Parker, & M. Stone (Producers). On *South Park*. New York: Comedy Central.

Pary, R., Lewis, S., Matuschka, P., Rudzinskiy, P., Safi, M., & Lippman, S. (2002). Attention deficit disorder in adults. *Annals of Clinical Psychiatry, 14*, 105-111.

Peterson, K., McDonagh, M. S., & Rongwei, F. (2008). Comparative benefits and harms of competing medications for adults with attention-deficit hyperactivity disorder: a systematic review and indirect comparison meta-analysis. *Psychopharmacology, 197*(1), 1-11.

Piotrowski, C., Belter, R. W., & Keller, J. W. (1998). The impact of "managed care" on the practice of psychological testing: preliminary findings. *Journal of Personality Assessment, 70*(3), 441-447.

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

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*(1), 33-65.

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status*. San Antonio, TX: The Psychological Corporation.

Rawling, P., & Brooks, N. (1990). Simulation Index: A method for detecting factitious errors on the WAIS--R and WMS. *Neuropsychology, 4*(4), 223-238.

Rehabilitation Act of 1973, 29 U.S.C. § 701 *et seq*.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives of Psychology, 28*, 286-340.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12*(1), 29-40.

Reynolds, C. R. (1998). Common sense, clinicians, and actuarialism in the detection of malingering during head injury litigation. In *Detection of malingering during head injury litigation.* (pp. 261-286). New York, NY, US: Plenum Press.

Rogers, R. (1997). *Clinical assessment of malingering and deception (2nd ed.)*. New York, NY: The Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms professional manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior, 22*(4), 353-367.

Rogers, R., Sewell, K. W., Cruise, K. R., Wang, E. W., & Ustad, K. L. (1998). The PAI and feigning: A cautionary note on its use in forensic-correctional settings. *Assessment, 5*(4), 399-405.

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment, 67*(3), 629.

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*(4), 349-359.

Saklofske, D. H., Tulsky, D. S., Wilkins, C., & Weiss, L. G. (2003). Canadian WISC-III directional base rates of score discrepancies by ability level. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 35*(3), 210-218.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213-223.

Slick, D. J., Hinkin, C. H., van Gorp, W. G., & Satz, P. (2001). Base rate of WMS-R Malingering Index in a sample of non-compensation-seeking men infected with HIV-1. *Applied Neuropsychology, 8*(3), 185-189.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18*(6), 911-922.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545-561.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: a survey of experts' practices. *Archives of Clinical Neuropsychology, 19*(4), 465-473.

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

Strauss, E., Slick, D. J., Levy-Bencheton, J., Hunter, M., MacDonald, S. W. S., & Hultsch, D. F. (2002). Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology, 17*(5), 423-444.

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*(4), 500-514.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in Attention-Deficit Hyperactivity Disorder and Learning Disorder assessments. *Applied Neuropsychology, 14*(3), 189-207.

Sumanti, M., Boone, K. B., Savodnik, I., & Gorsuch, R. (2006). Noncredible psychiatric and cognitive symptoms in a workers' compensation "stress" claim sample. *The Clinical Neuropsychologist, 20*(4), 754-765.

Sweet, J. J., Peck, E. A., Abramowitz, C., & Etzweiler, S. (2002). National Academy of Neuropsychology/Division 40 of the American Psychological Association practice survey of clinical neuropsychology in the United States, Part I: practitioner and practice characteristics, professional activities, and time requirements. *The Clinical Neuropsychologist, 16*(2), 109-127.

Szasz, T. S. (1956). Malingering: diagnosis or social condemnation? Analysis of the meaning of diagnosis in the light of some interrelations of social structure, value judgment, and the physician's role. *A. M. A. Archives Of Neurology And Psychiatry, 76*(4), 432-443.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI–2 Restructured Clinical scales: Development, validation, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Thornicroft, G., Rose, D., Kassam, A., & Sartorius, N. (2007). Stigma: ignorance, prejudice or discrimination? *The British Journal of Psychiatry, 190*, 192-193.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260-268.

Trueblood, W. (1994). Qualitative and quantitative characteristics of malingered and other invalid WAIS--R and clinical memory data. *Journal of Clinical and Experimental Neuropsychology, 16*(4), 597-607.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology, 15*(4), 578-590.

USDE. (2005). Office of Special Education and Rehabilitative Services, Annual Report to Congress on the Implementation of The Individuals with Disabilities Education Act, selected years, 1977 through 2005; and Individuals with Disabilities Education Act (IDEA) database. April 22, 2008, from http://www.ideadata.org/PartBdata.asp.

Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., et al. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*(1), 37-48.

Victor, T. L., Boone, K. B., Serpa, G. J., Buehler, J., & Zeigler, E. A. (In Press). Interpreting the meaning of multiple symptom validity test failure. *Clinical Neuropsychologist*.

Wade, J. B., & Hart, R. P. (2002). Attention and the stages of pain processing. *Pain Medicine (Malden, Mass.), 3*(1), 30-38.

Wang, E. W., Rogers, R., Giles, C. L., Diamond, P. M., Herrington-Wang, L. E., & Taylor, E. R. A pilot study of the personality assessment inventory (PAI) in corrections: assessment of

malingering, suicide risk, and aggression in male inmates. *Behavioral Sciences & the Law, 15*(4), 469-482.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Test-Revised: Manual*. New York, NY: The Psychological Corporation.

Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. New York, NY: The Psychological Corporation.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale (3rd ed.)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale (3rd ed.)*. San Antonio, TX: The Psychological Corporation.

White, B. P., Becker-Blease, K. A., & Grace-Bishop, K. (2006). Stimulant medication use, misuse, and abuse in an undergraduate and graduate student sample. *Journal of American College Health, 54*(5), 261-268.

Wilde, N. J., Strauss, E., & Tulsky, D. S. (2004). Memory Span on the Wechsler Scales. *Journal of Clinical & Experimental Neuropsychology, 26*(4), 539-549.

Williams, R. W., & Carlin, M. (1999). Malingering on the WAIS--R among disability claimants and applicants for vocational assistance. *American Journal of Forensic Psychology, 17*(1), 35-45.

Willison, J., & Tombaugh, T. N. (2006). Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology, 21*(1), 41-52.

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

Yelin, E. (1986). The myth of malingering: why individuals withdraw from work in the presence of illness. *The Milbank Quarterly, 64*(4), 622-649.

Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2*(3), 279-283.

Youngjohn, J. R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *The Clinical Neuropsychologist, 9*(2), 112-123.

Zhu, J., & Tulsky, D. S. (2000). Co-norming the WAIS-III and WMS-III: Is there a test-order effect on IQ and memory scores? *The Clinical Neuropsychologist, 14*(4), 461.

Vita

Russell Pella attended undergraduate study at the University of Nebraska at Kearney in Kearney, Nebraska, where he majored in criminal justice and psychology. He then completed a master's degree in clinical psychology from New Mexico Highlands University in Las Vegas, New Mexico, under the direction of Steven Fox and Gerald Russell. During his doctoral studies at Louisiana State University in Baton Rouge, Louisiana, he had the pleasure of being supervised, mentored, and befriended by Wm. Drew Gouvier who taught lessons spanning the academic and applied settings. He was accepted to perform his American Psychological Association accredited clinical psychology internship at the University of Oklahoma Health Sciences Center in order to specialize in neuropsychology where he primarily studied under the mentorship of Russell Adams and James Scott. He was also accepted to perform a two-year clinical neuropsychology postdoctoral fellowship at the South Texas Veterans Health Care System, which is accredited by the American Psychological Association and Association of Postdoctoral Programs in Clinical Neuropsychology.