2011

Simulated subnormal performance on the Stanford Binet-V: an exploratory investigation of the Stanford Binet rarely missed items index

Mandi Wilkes Musso
Louisiana State University and Agricultural and Mechanical College, mmusso4@tigers.lsu.edu

SIMULATED SUBNORMAL PERFORMANCE ON THE STANFORD BINET-V: AN EXPLORATORY INVESTIGATION OF THE STANFORD BINET RARELY MISSED ITEMS INDEX

A Thesis
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
Master of Arts

in

The Department of Psychology

by
Mandi W. Musso
B.S., University of Louisiana at Lafayette, May 2006
May, 2011

ACKNOWLEDGEMENTS

Dr. Gale Roid and Riverside Publishing Company are gratefully acknowledged for providing standardization data from the Stanford Binet, Fifth Edition. These data were used in the development and initial validation of the Stanford Binet-5 Rarely Missed Items Index (SBRMI). I would also like to thank my thesis committee, Dr. Alex Cohen, Dr. Mike Hawkins, Dr. Glenn Jones, and my advisor, Dr. Wm. Drew Gouvier.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of this study was to derive an embedded validity index of effort for the Stanford-Binet Intelligence Scales-Fifth Edition (SB5; Roid, 2003a) and to validate this index using an analog sample of individuals feigning mild mental retardation (MR). Of the data provided by Dr. Roid, 307 healthy individuals aged 18-35 with full scale intelligence quotients (FSIQ) greater than 70 were included in this study (n = 307) as well as 31 individuals with FSIQ scores in the MR range. Also, a sample of 108 undergraduate students at Louisiana State University was asked to participate in this study in exchange for extra credit in an undergraduate psychology class. One group, analog malingerers, was instructed to feign mild MR and respond to the material in a manner consistent with that population while those assigned to the control group were asked to perform to the best of their ability. Frequencies of incorrect responses were calculated for each item and items retained were missed significantly more frequently, $p < .001$, by individuals asked to feign MR. Twenty-one items were retained and compose the Stanford-Binet, Fifth Edition Rarely Missed Items Index (SBRMI). Logistic regression analyses indicated that the SBRMI was a significant predictor of malingered MR, and ROC curve suggested that a cutoff score of 17.50 yielded a sensitivity of 70.5% and a specificity of 100%. Overall, it appears that the SBRMI is a clinically useful tool for detecting malingered mental retardation.

INTRODUCTION

Accuracy of diagnostic decisions is constrained by the skill of the diagnostician, the operating characteristics of the tests used, and the effort exerted by the client during the evaluation. It has, for decades, been widely accepted that a test must possess acceptable psychometric properties; however, the necessity of measuring effort exerted by clients had been lagging in the literature until the 1990s. Individuals might distort responses or fabricate symptoms for multiple reasons. In conversion disorder, for example, an individual may involuntarily and without awareness exaggerate or feign symptoms in response to intrapsychic or extrapsychic stressors. Most symptoms in conversion disorder are related to sensory and motor functions and are referred to as “pseudoneurological” by the Diagnostic and Statistical Manual of Mental Disorders, Fourth edition revised (DSM-IV-TR, American Psychological Association, 2000; refer to Appendix A for list of Abbreviations). In contrast, Factitious Disorder involves intentional exaggeration or feigning of symptoms in the presence of psychological incentives, specifically, with the motivation being a desire to assume a “sick role”. Factitious disorder is differentiated from malingering by the absence of external incentives (i.e. monetary gain, avoiding legal responsibility; APA, 2000).

In neuropsychological literature, the lion’s share of research and attention on insufficient effort and symptom fabrication has been paid to malingering. Malingering, while not an actual psychological disorder, is defined as the intentional fabrication of symptoms in order to obtain an external incentive. The definition and assumed models of malingering have been revised many times over the years and have evolved with increasing interest in this topic. Rogers (1990a) called for psychologists to turn their attention from an earlier conception of malingering labeled the Puritanical Model, which focused on the moral “badness” of malingering, toward and empirically based detection model that focused on research designs using either known groups (individuals known to be malingering versus individuals with known deficits) or simulation groups (assigning subjects to feign impairment or control groups). He promoted a movement that encouraged empiricism over morality and lead to the improved science of evidence based practice. Rogers (1990b) proposed the following two tentative criteria for malingering, including (A) a symptom endorsement cluster: endorsement of high numbers of rare symptoms (not typically seen in actual patients), blatant symptoms (obviously severe psychopathology), endorsement of such high numbers of symptoms that it is improbable based solely on the number of symptoms, or endorsement of symptoms which are absurd in alert and oriented individuals and (B) a convergent/divergent validity cluster: corroboration of feigned impairment based on reports from reliable observers (i.e. family members), discrepancies between self-reported prior episodes and documented history, or evidence of feigning on standardized measures.

Greiffenstein, Baker, and Gola (1994) offered a research study with operational definitions of malingering. These authors classified individuals as probable malingerers if they met two of the following criteria: two or more “severe” impairment ratings on neuropsychological tests, improbable symptom history which is inconsistent with records or surveillance data, “total disability” in work or social roles after one year, or claims of retrograde memory loss. This study represents progress in the literature by furthering the search toward an operationally defined and empirically based study of malingering.

Slick, Sherman, and Iverson’s (1999) paper, drawing from Rogers (1990b) and Greiffenstein et al (1994), revolutionized the study of malingering by providing working operational definitions of malingering categorized by degree of clinician certainty and based on well-defined and objective diagnostic criteria (Table 1). They proposed that malingered neurocognitive dysfunction (MND) exhibited during psychological or neuropsychological evaluations could be categorized based on degree of diagnostic certainty as possible MND, probable MND, and definite MND. In order to be diagnosed as MND, an individual must present with substantial external incentive (Criteria A) including but not limited to monetary gain and avoiding or reducing criminal prosecution. In addition, evidence must be obtained from objective testing (Criteria B) including either definite negative response bias defined as below chance performance on a forced-choice test or probable response bias defined as performance consistent with feigning on well validated effort indices, or discrepancies between test data and known levels of neurocognitive functioning, behavioral observations, reliable informant's report of daily functioning, and known background history. Evidence from self-report (Criteria C) can also be examined in diagnosing MND including discrepancies between self-reported history and documented history; or self-reported symptoms and known patterns of brain functioning, behavioral observations, information from reliable informants, as well as evidence of exaggerated or manufactured psychological functioning. Finally, observed data must not fully be accounted for by psychological, medical, developmental or neurological deficits (Criteria D). In order for an individual to be diagnosed as definite MND, they must have a substantial external incentive (Criterion A), show definite response bias (Criteria B; performance below chance at the $p < .05$ on a binary forced choice test), and these deficits must meet criteria D. Probable MND is diagnosed in individuals meeting Criteria A and D in addition two or more types of evidence (which fall short of below chance performance) from testing (Criteria B) or one type of evidence from testing and one type from self-report (Criteria B and Criteria C, respectively).

Table 1. Malingering criteria proposed by Slick et al. (1999)

Criteria for malingering
Criterion A: One identifiable, substantial external incentive
Criterion B: Evidence from test data
1. Definite Response Bias: below chance ($p < .05$) performance on at least one FCT
2. Probable Response Bias: performance consistent with malingering on at least one validated measure or index of malingering
3. Discrepancy between performance and known patterns of functioning: must be consistent with known pattern of exaggeration
4. Discrepancy between test data and observed behavior: data from two or more tests within one domain are discrepant with observed functioning
5. Discrepancy between test data and reliable collateral reports: data from tests of at least one domain are discrepant with day-to-day functioning
6. Discrepancy between test data and background/history: data from two or more tests of a domain are discrepant with reported neurological history
Criterion C: Evidence from self-report
1. Discrepancy between self-report and documented history: consistent with attempt to exaggerate deficits
2. Discrepancy between self-report and known patterns of brain functioning: unlikely in number, severity, or pattern or inconsistent with known functioning
3. Discrepancy between self-report and observed behavior: self-report is discrepant with observed functioning
4. Discrepancy between self-report and reliable collateral reports: self-reported symptoms are discrepant with day-to-day functioning reported by reliable informants
5. Evidence of exaggerated or fabricated dysfunction: evidence from self-report and test data suggest exaggeration or malingering
Criterion D: Behavior not fully accounted for by psychological, neurological, or developmental factors
Classification Criteria
Definite MND: Criteria A, B1, and D
Probable MND: Criteria A, B2-6 OR C1-5, and D
Possible MND: either Criteria A, C1-5, and D OR criteria for Probable MND except for criteria D

Possible MND is diagnosed by either Criterion A, one type of evidence from testing self-report (Criteria C), and Criterion D, or classification of Definite or Probable MND met except for Criterion D. It should be noted that in the latter instance, alternative etiologies cannot be ruled out without further information.

The *DSM-IV-TR* currently includes malingering as a condition that should be the focus of additional clinical attention (V65.2). The *DSM-IV-TR* suggests that malingering should be suspected in a medicolegal context, if an individual exhibits a lack of cooperation during an evaluation, performance that is markedly different from what would be expected based on person's disability, or has a diagnosis of Antisocial Personality Disorder. These criteria are current guidelines which fall short of actual diagnostic criteria and might be viewed as persisting vestiges of the Puritantical Model of Malingering which Rogers (1990a) warned should be set aside in light of a more objective model.

Detection of Malingering

It is widely held that the use of only one validity measure provides insufficient information to determine whether or not a person is malingering in a psychological evaluation (Larrabee, 2007). Researchers recommend using multiple validity measures to detect malingering, including symptom validity tests (SVTs), also known as forced-choice tests (FCTs), embedded indices, and other tests with established sensitivity and specificity (Iverson and Binder, 2000). Larrabee (2007) reported a "significant increase in the probability of malingering with aggregation of scores that fall in the malingered range of performance" (p. 9).

Diagnostic Statistics

Diagnostic statistics provide information about the accuracy of a specific test in classifying individuals with and without specified disorders (Larrabee & Berry, 2007), and are critical in evaluating the accuracy and applicability of tests of MND (Bianchini, Mathias, & Greve, 2001). Important diagnostic statistics in determining classification accuracy include base rates, sensitivity, specificity, positive predictive power (PPP), negative predictive power (NPP) and likelihood ratio (Gouvier, Hayes, & Smiroldo, 1998). Base rates are the prevalence of a given condition in a population, and they exert enormous influence on diagnostic classification (Gouvier, 2001). Sensitivity reflects the probability that an individual with a particular disorder will be diagnosed, and it is calculated as the ratio of true positives divided by the sum of true positives and false negatives. Specificity refers to the probability that a person without a specified disorder obtains a negative test result and is calculated by dividing true negatives by the sum of true negatives and false positives (Larrabee & Berry, 2007). Positive predictive power is the probability of a disorder being diagnosed given a positive test finding, and NPP is the probability of an individual without a given disorder being classified as such given a negative test finding. Both PPP and NPP are influenced by the base rate of a given disorder and will change as the base rate changes.

Stand-Alone Tests of Malingering

Forced-choice Effort Tests (FCT). In FCTs, respondents are presented with a two-choice response format in which one choice is a target, previously presented, and the other is a foil. Statistically, the probability of answering correctly depending on chance would equal approximately 50%. Originally, FCTs were designed to examine responses below chance levels based on the assumption that guessing would yield approximately chance-level scores.

In this paradigm, scores significantly below chance confirm the respondent's ability to 1) differentiate among the response targets and 2) choose the wrong response for recording. Significantly below chance and chance level responding are rather rare, however. Further research suggests that cutoff scores are often useful in detecting individuals who are exhibiting suboptimal performance because the individual need not generate a below-chance performance level to fail. If an individual performs below empirically derived and well validated cutoff scores for neurologically impaired or intellectually disabled individuals, it can be assumed that they are either not providing sufficient effort by responding carelessly or that they are intentionally distorting their responses.

The Word Memory Test (WMT) is a computer administered test of verbal learning and memory which presents six scales of performance including two forced choice measures of recognition memory. According to its authors, individuals who score below 82% on immediate recognition or delayed recognition trials should be suspected of poor effort (Green, Allen, & Astner, 1996). It has been determined that the false positive rate of the WMT is very low (Green, Flaro, & Courtney, 2009). In one study, 64 individuals with moderate to severe traumatic brain injury obtained mean scores greater than 90% (Green Iverson, & Allen, 1999). Aforementioned studies have found that effort influences WMT scores and accounts for a greater proportion of variance in scores on the WMT than neurological dysfunction or traumatic brain injury. These studies provide further evidence for the utility of the WMT in detection of poor effort or intentional distortion of responses (Green et al., 1999; Green et al., 2001; Flaro, Green, & Robertson, 2007).

Another popular symptom validity test is the Test of Memory Malingering (TOMM; Tombaugh, 1996). Rees et al (1998) conducted an extensive validation study using five experimental groups and found that the TOMM, with a cutoff score of 45, is a useful tool in the detection of malingering. The TOMM has demonstrated the ability to accurately distinguish individuals with traumatic brain injury from malingerers (Greve, Bianchini, & Doane, 2006; Harber & Fichtenberg, 2006). Using WMT failure as the operational criterion, O'Bryant and Lucas (2006) calculated .98 PPP and .78 NPP of the TOMM using the data of Gervais et al (2004). In a heterogeneous sample of non-litigating German patients with "diagnosable cerebral disease" (n = 48) and Alzheimer's Disease (n = 20), the TOMM was not sensitive to neurological impairment but proved relatively robust against false positive errors (Merten, Bossink, & Schmand, 2007).

Comparing TOMM and WMT. The Word Memory Test (Green, 2003) is considered by some to be the "gold standard" of FCT's (O'Bryant & Lucas, 2006). It has demonstrated greater sensitivity than the TOMM (Gervais, Rholing, Green, & Ford, 2004; Green, 2003; Green, 2007). In one study where 694 participants were administered both the TOMM and the WMT, 111 cases passed the TOMM but failed the WMT where only three cases failed the TOMM but passed the WMT (Green, 2003). A comparison of TOMM and WMT performance of 519 adults who were either personal-injury litigants, disability seeking, or involved in workers compensation claims found that the TOMM misclassified 69% of individuals who failed one other SVT. In this study, the WMT had the most failures (32%) compared to the TOMM (11%) indicating that it was more sensitive to inadequate effort (Gervais et al., 2004). It is unclear whether the TOMM is easily identified as a malingering measure or whether the WMT is more difficult to deceive.

While Green (2007) reported significantly higher failure rates of the WMT compared to the TOMM suggestive of better sensitivity of the WMT, some researchers have questioned his methodology stating that he compared failure of only Trail 2 of the TOMM to failure of one of several measures on the WMT (Greiffenstein, Greve, Bianchini, & Baker, 2008). In evaluating individual's performance using Trial 1, Trial 2, and the retention trial of the TOMM, these investigators determined that the WMT did show slightly better sensitivity but also classified more individuals with severe injuries as malingers than did the TOMM suggesting possibly poorer specificity. In a study of 1032 individuals evaluated in a clinical psychology/neuropsychology practice, more individuals performed below chance on the WMT (IR, DR, and IR + DR indices) than the TOMM (Trial 1, Trail 2, and Retention Trial); however, the TOMM was as accurate at identifying suboptimal performance as the WMT when empirically derived cutoff scores (< 45) are used (Greve, Binder, & Bianchini, 2009). A study of non-litigating individuals with severe cognitive impairment and obvious symptoms (either Alzheimer's disease or other diagnosed cerebral disease) found that the TOMM was failed least often than the WMT (Merten, Bossink, & Schmand, 2007). For individuals with Alzheimer's disease, almost all patients failed the WMT where 95% passed the TOMM; in addition, 17% of patients with other cerebral disease failed the TOMM where 50% failed the WMT IR index and 42% failed the WMT DR index. Merten et al (2007) suggest that some SVTs may not be sensitive to severe cognitive impairment and determination of effort should not rest solely SVT scores.

Floor Effect Strategies. Malingering indices that rely on the "floor effect" have demonstrated some utility in the detection of malingering. Normative floor effect strategies compare an individual's performance to performance of persons with known impairment (Frederick, 2000). The TOMM and WMT are both FTCs that employ normative floor effect strategies.

Additional Stand-alone Tests of Malingering. Popular stand alone tests used in the field of clinical neuropsychology include the Dot Counting Test (Boone et al., 2002; Rey, 1941), Rey's Fifteen-Item Test, also known as the Memory for Fifteen-Item Test (Rey, 1964; Lezak, 1995), the Portland Digit Recognition Test (Binder & Willis, 1991), the M-Test (Beaber, Marston, Michelli, & Mills, 1985), the Victoria Symptom Validity Test which was originally designed by Hiscock and Hiscock (1989) and then modified by Slick et al. (1994, 1996, 1997), and the Computerized Assessment of Response Bias (Conder, Allen, & Cox, 1992). Stand alone tests used in detecting feign psychological impairments include the Miller Forensic Assessment of Symptoms Test (Miller, 2001), and the Structured Interview of Reported Symptoms (Rogers, Bagby, & Dickens, 1992).

Embedded Validity Indices

While standalone measures are designed specifically to detect malingering and poor effort, and usually have high sensitivity and specificity, embedded indices are continuing to grow in popularity. Embedded indices have utility because they add to the armamentarium that psychologists and neuropsychologists use in the detection of malingering without extending important resources such as time and money with the addition of tests to the battery. In addition, embedded indices can provide valid effort indicators for many of the specific tests actually used in the determination of neuropsychological deficits. They are also less likely to be coached than symptom validity tests (Heinley et al. 2005). The use of embedded effort indices along with stand-alone measures represents the practice of using a multi-trait multi-method approach (Campell & Fiske, 1959) that has underpinned modern approaches to psychological and neuropsychological assessment over the past several decades. Many embedded validity indicators are derived from the same techniques reflected in the stand-alone tests mentioned previously, but many other techniques are employed as well in the derivation of embedded validity indices.

Millis and Volinsky (2001) found that only a minority of malingerers perform below chance on forced choice measures and that the format is easily recognized so that more sophisticated malingerers are able to identify FCTs and avoid detection of insufficient effort. For this reason, researchers sometimes use other methods that capitalize on alternate signs of malingering including: performance differences on parallel tests, significantly better anterograde than retrograde memory in persons with traumatic brain injury, performance which is similar or better on easier rather than difficult versions of a test, the absence of a learning curve (lack of improvement where expected; Hall & Pritchard, 1996), and the preservation of implicit learning and memory in persons with amnestic disorders. Pattern analyses are important in understanding how malingerers differ from persons with actual impairment (Iverson and Binder, 2000). It is essential for clinicians to understand how persons from the general population perform on tests, how nonlitigants with similar conditions perform, and how persons classified as probable or definite malingerers perform on given tests. By utilizing this information in the interpretation of an individual's test scores, clinicians will be better able to determine the likelihood an individual is malingering. Some examples of the embedded measures approach to malingering detection are presented below:

Rarely Missed Items Index. The Rarely Missed Items index was proposed by Killgore and DellaPietra (2000). They developed this index from the Logical Memory Delayed Recognition subtest of the WMS-III. In the Logical Memory Delayed Recognition subtest, individuals are asked to answer thirty "yes" and "no" questions about Logical

Memory stories they were previously read. Killgore and DellaPietra (2000) predicted that individuals who had never heard the Logical Memory stories would answer some questions correctly at above chance levels. They established rarely missed items by having students who were naïve to the test complete the Logical Memory Delayed Recognition subtest and found six questions were answered correctly by 70-80% of students. In a second experiment, the authors asked analog malingerers and patients with neuropsychological impairment to complete the Logical Memory Delayed Recognition subtest after hearing the stories. Using weighted combinations of the six items, the examiners determined a cutoff score of 136 or below exhibited a sensitivity of 97% and specificity of 100%. In this study, analog malingers were more likely to choose incorrect answers on the six questions determined to have a low probability of being missed.

Subsequent research on the Rarely Missed Items index of the WMS-III has been conducted in various populations and yields mixed results. Marshall and Happe (2007) determined that individuals with mild MR tend to answer "yes" when they don't know the correct answer. Because the correct answer for each of the six items on the Rarely Missed Items is "yes", individuals with mild MR tend to score above the cutoff for malingering. However, a sample of college students (*n* = 100) who were naïve to the logical memory stories, showed response biases only on half of the Rarely Missed Items (Swihart, Harris, & Hatcher, 2008); in addition, the authors found that the Rarely Missed Items had lower sensitivity (.18), specificity (.87), positive predictive power (.45) and negative predictive power (.64) in analog malingerers given multiple tests and incentives to successfully feign head injury compared to sensitivity and specificity reported by Killgore and DellaPietra (2000). The authors propose that when individuals are given multiple tests and instructed to feign head injury, they tend to be more conservative and choose not to distort responses on the WMS-III Rarely Missed Items. Research on the utility of the Rarely Missed Items with forensic criminal populations appears less promising. Lange, Sullivan, and Anderson (2005) investigated the utility of the RMI index in 158 litigants and 78 nonlitigants with mixed diagnoses. Overall, the authors obtained a sensitivity rate of .25, specificity rates ranging from .91-.95, positive predictive power ranging from .50-.71, and negative predictive power ranging from .68-.83. In addition, the authors reported relatively low rates of positive scores for suspected exaggerators and borderline exaggerators indicating that multiple individuals who were expected of malingering were not detected by the Rarely Missed Items. D'Amato and Denney (2008) investigated the utility of the Rarely Missed Items index in a criminal population and concluded that it demonstrated poor diagnostic utility in distinguishing suspected malingers from valid performers with sensitivity of .33, specificity of .83, (with base rate of .23) PPP of .38, and NPP of .81.

Pattern Analysis. Malingerers may be able to successfully simulate the profile of intellectual deficits seen in individuals with actual head injury (Heaton, Smith, Lehman, & Vogt, 1978). Therefore, it becomes necessary to evaluate patterns of performance on individual subtests in neuropsychological assessment in order to identify malingerers. Heaton et al. (1978) first asserted that malingerers could be identified based on patterns of performance on psychological and neuropsychological tests, specifically, the MMPI, WAIS, and Halstead Reitan battery. Secondly, and reaffirming Meehl's (1956) findings, they confirmed that a methodology based on multivariate statistical analyses can almost always better detect the diagnosis of interest than clinical judgment. In this study, they used two discriminate function analyses, one for the MMPI and another for the neuropsychological variables. The cutoff scores derived from the neuropsychological data correctly classified head injured (*n* = 16) vs. malingering (*n* = 20) subjects with 100% accuracy, while the MMPI misclassified only one individual in each group. Upon cross validation, in a sample of clinic referrals either known to be involved in litigation and suspected of exaggeration (*n* = 42) or without external incentive and exhibited good effort (*n* = 42), the discriminate functions identified more individuals in the suspected malingering group (64.3%) compared to the good effort group (26.2%; Heaton et al., 1978). Although the sample size was small and further research could only partially replicate the findings (Thompson & Cullum, 1991), this study laid the foundation for future research in pattern analysis profiles and malingering.

Discriminate function analysis performed on the WAIS-R defined profiles that were characteristic of analog malingerers (*n* = 67) compared to non-ligitating individuals who had a documented head injury (*n* = 67) with a sensitivity 70% of and specificity of 78% (Mittenberg, Theroux-Fichera, Zielinski, & Heilbronner, 1995). In the same article, the authors determined that significantly lower Digit Span compared to Vocabulary scores reliably classified poor effort in 71% of individuals asked to simulate head injury compared to non-litigating individuals with documented traumatic brain injuries. Additional research determined that the Vocabulary-Digit Span index is useful in detecting poor effort in a sample of litigating individuals with mild traumatic brain injury (classified as probable malingerers) compared to non-litigating individuals with traumatic brain injury (Greve et al., 2003). Furthermore, in a study designed to investigate the Vocabulary-Digit Span difference score and the discriminate function proposed by Mittenberg et al. (1995) individuals with moderate to severe head injuries admitted to a rehabilitation hospital were compared with litigating individuals with mild traumatic brain injury (Millis, Ross, & Ricker, 1998). Millis et al. (1998) found that Mittenberg's discriminate function analysis (90% correct classification) was more effective than the Vocabulary-Digit Span discrepancy score (79% correct classification) at detecting malingered neurocognitive dysfunction.

Psychological and Personality Measures. Measures of personality and psychological functioning often include indexes that measure effort and attempts at response distortion. For example, the Minnesota Multiphasic Personality Inventory, Second Edition (MMPI-II; Butcher et al., 1989) F, FBS, Fb, F-K, Fp scales can be used to identify attempts to "fake bad" while the L and K scales measure endorsement of socially desirable responding and impression management. Similarly, the Personality Assessment Inventory (Morey, 1991) has Positive and Negative Impression Management scales that examine individual's attempts to portray themselves in more favorable or more unfavorable terms, respectively. The Personality Assessment Inventory also has scales that detect responses to items answered infrequently by individuals with actual psychiatric disorders and inconsistent response styles (i.e. answering similar questions in different fashions).

Reliable Digit Span. The Reliable Digit Span index was established by Greiffenstein, Baker and Gola (1994). The score is determined by first calculating the longest trial in which both strings were correctly recalled for both the forward and backward tests then summing the lengths of the longest string recited accurately for each. The recommended cutoff score was seven for the sum of the longest digits forward and longest digits backward. Scores below seven are indicative of inadequate effort.

Prevalence of Malingering

A survey of American Board of Clinical Neuropsychology members found that neuropsychologists in private or group practice estimated base rates of probable malingering are about 30% in civil cases and 21% in criminal cases (Mittenberg, Patton, Canyock, & Condit, 2002). This survey also concluded that in criminal forensic cases, base rates for feigned impairment were higher among patients referred by the prosecution versus those evaluated in the context of their own defense (Mittenberg et al., 2002). An important, but unanswered question remains: “what proportion of the people who are classified as “malingerers” actually have comorbid brain dysfunction as well?”. In another survey, 24 expert neuropsychologists from the North America were asked to estimate the base rates of malingering in their forensic practices. Two-thirds of the experts estimated base rates of definite malingering over 10% and one-third of these experts estimated 20% or higher base rate of definite malingering (Slick, Tan, Strauss, and Hultsch, 2004).

Medicolegal Setting. Medicolegal settings tend to be hotbeds of malingering because successfully feigned neurocognitive impairment may result in substantial financial gain. Mittenberg et al. (2002) asserted that 38.5% (41.24% after adjusting for referral source) of personal injury litigants who claimed that they’d sustained mild head injuries were identified as probable malingerers (Mittenberg et al., 2002). An examination of 11 studies providing information on the base rates of malingering found that 40% performed in a manner consistent with malingering (Larrabee, 2003); however criteria for each study varied and no clear unitary definition of malingering was provided.

Bianchini, Curtis, and Greve (2006) investigated the prevalence of malingering among individuals with traumatic brain injury who were either not involved in litigation, were seeking compensation from the state, or were pursuing federally subsidized workers compensation. They found a “dose-response” relationship in that the probability of malingering increased as the level of potential compensation payoff increased (Bianchini, Curtis, & Greve, 2006). In this study, individuals involved in state workers compensation litigation met Slick et al. (1999) criteria more frequently (*n* = 184; 17.7%) than individuals who were non-litigating (*n* = 94; 0%), and individuals involved in federal workers compensation cases (higher awards) had the most invalid responses of all (*n* = 54; 33%).

A study of 904 Canadian participants who were involved in either worker's compensation claims (*n* = 376), medical disability claims (*n* = 317), or personal injury litigation (*n* = 196) found that the effort put forth by participants during a neuropsychological evaluation explained 53% of the variance in test performance, whereas education and age accounted for 11% and 4%, respectively (Green, Rholing, Lees-Haley, & Allen, 2001). Individuals with mild head injury tend to score significantly lower on SVTs, specifically the WMT, compared to those with moderate-to-severe head injuries, suggesting that individuals with mild head injury put forth less effort or tend to distort responses more than individuals with more severe documented head trauma (Green, Iverson, & Allen, 1999; Green et al., 2001; Flaro, Green, & Robertson, 2007). Of 577 individuals claiming mild head injury who were involved in worker's compensation, disability, or personal injury claims, 40% failed the WMT (Flaro et al., 2007). Given that mild head injuries are far more prevalent than severe head injuries, the problem of malingered mild head trauma is enormous with direct financial cost estimated at over five billion dollars a year during the decade of the 1990's (Gouvier, Lees-Haley, & Hayes-Hammer, 2003).

Individuals with mild head injury frequently exhibit significantly poorer performance compared to persons with severe or moderate head injuries in a medicolegal setting (Greiffenstein & Baker, 2006). Lees-Haley and Brown (1993) found that personal injury litigants both with and without head injury have higher base rates of current self-reported neuropsychological symptoms than controls. An investigation into self-reported ratings of perceived current and past functioning among a sample of primarily head injured litigants versus non-litigating individuals found that plaintiffs scored their premorbid functioning as being significantly higher and their current functioning as being significantly lower compared to non-litigating individuals (Lees-Haley et al, 2007). It should be noted that in this study, non-litigating individuals were asked to rate prior functioning three years ago whereas plaintiffs were instructed to rate premorbid functioning. Further research indicates that even non-litigating individuals who sustain a closed head injury have more inaccuracy in retrospective recall of neuropsychological and psychological functioning compared to individuals who sustain back-injuries or significant psychosocial stressors (Hilsabeck, Gouvier, & Bolter, 1998). In the same study, the authors found that individuals who sustain physical injuries tend to under report pre-injury symptoms even when they have no external incentive to do so (Hilsabeck, Gouvier, & Bolter, 1998). The authors concluded that Lees-Haley et al (1997) findings are premature and were driven more by injury status than litigation status.

Criminal Forensic Setting. Decisions about competence and mental capacity of alleged criminals may result in reduced placement outcomes, sentences commonly perceived as “easier”, such as in a mental institution rather than a correctional facility, or avoidance of the death penalty (Atkins v. Virginia, 2002). In competency to stand trial cases, pretrial individuals are evaluated for the presence of either severe and uncontrollable psychological symptoms or sufficiently limited knowledge of the judicial system to render the individual incapable of participating in his or her defense such persons are typically remanded to a mental facility for competency restoration training. Individuals adjudicated not guilty by reason of insanity are typically hospitalized in high security institutions and eventually released into the community if and when deemed suitable for release. In one study of 105 pretrial competency to stand trial cases, 54.29% were classified as definite or probable malingerers with an additional 26.7% classified as possible malingerers (Ardolf, Denney, and Houston, 2007). By comparison the rate of malingered neurocognitive dysfunction in general medical cases is estimated at 7-12% (Mittenberg, Patton, Canyock, & Condit, 2002). The base rates of malingering have been found to be lower in not guilty by reason of insanity compared to competency to stand trial cases (Weinborn, 2003).

In 2002, a ruling in the case of Atkins v. Virginia determined that convicted criminals with MR would no longer be eligible for the death penalty. Because MR is associated with diminished capacity for reasoning and appreciation, it was deemed cruel and unusual punishment and a violation of the Eighth Amendment. According to the *DSM-IV-TR*, in order for an individual to be considered MR, he or she must display an IQ less than or equal to 70, deficits in at least two adaptive functioning domains, and recognized disability before the age of eighteen. During the Atkins trial, the prosecution's expert witness testified about the defendant's IQ using information obtained from two interviews with the defendant, interviews with correctional officers, and a review of school records. During the appeal, Justice Hassel of Virginia's Supreme Court stated that the opinion of the prosecution's expert witness, who claimed the defendant was not mentally retarded, was "incredulous as a matter of law" (Atkins, 2002, Dissent Justice Hassel, III, as cited in Everington & Olley, 2008). Justice Hassel's statement exemplifies the necessity for standardized, objective methodology for the determination of MR. With the recognized utility of objective measures of IQ comes an added need for embedded validity indicators that provide information, not about the validity of the test itself, but about the level of overall effort exhibited by the individual being evaluated. Forensic cases frequently receive multiple evaluations, and the implications of practice effects can be profound. For example, in Atkins cases where an individual's IQ is at the upper margin of the MR range (close to 70), an increase in full scale IQ scores due to practice effects could literally make the difference between life and death. Additionally, in cases where individuals are adjudicated not guilty by reason of insanity, the possibility of malingering before the verdict or dissimulation afterwards is a concern. While one might not be able to "fake good" on intelligence tests without overt coaching, practice effects may greatly improve performance and this could mask the presence of actual deficits in individuals completing IMEs for medico legal or criminal forensic purposes.

Malingering in the Intellectually Disabled. While the detection of malingering in individuals with MR is of great potential importance, few researchers have examined the usefulness of neuropsychologists' current armamentarium of effort indices among members of this population. Misclassification of an individual with actual MR as malingering could potentially cost them their life. Research using individuals with no external incentive to malinger suggests that failure on effort tests is inversely correlated with full scale IQ; lower scores are associated with more effort test failures (Dean, Victor, Boone, & Arnold, 2008).

Utility of malingering indices varies greatly in MR populations. In one study of 21 adjudicated males with a confirmed diagnosis of mild MR who were administered the TOMM, only one participant scored below the cutoff (45) on Trial 2 with a mean score of 48.7 (Simon, 2007). In study of community volunteers with MR and community volunteers asked to feign MR, the TOMM showed a sensitivity of .56 and specificity of .96 (Graue et al, 2007). Additionally, Hale, Hayes, and Gouvier (1998) added the SIRS to the Memory for

Fifteen Items Test, Dot Counting Test, and M-Test, and obtained correct classification of 95% of MR individuals in a forensic setting as malingering or non-malingering. In a more comprehensive study of malingering indices in individuals with MR, Marshall and Happe (2007) determined that individuals with mild MR obtained scores above the empirically derived cutoffs on the Voc-Digits (98% passed), the Rarely Missed items (91% passed), and the California Verbal Learning Test-II (89% passed). Nonetheless, in the Marshall and Happe (2007) study, the Memory for Fifteen Items Test (21% passed), the Rey Dot Counting test (51% passed), and the Reliable Digit Span (31% passed) were measures that did not show adequate specificity in this population. In another study, individuals suspected of malingering MR actually performed better than those with MR and no external incentives on the Memory for Fifteen Items Test, Dot Counting Test, and M-Test (Hale, Hayes, & Gouvier, 1997). Graue et al. (2007) determined that the WAIS-III embedded validity indices were not able to discriminate true MR from feigned MR.

Hurley and Deal (2006) investigated the utility of the Memory for Fifteen Items Test, Dot Counting Test, TOMM, and Structured Interview of Reported Symptoms in individuals with no criminal history and with borderline intellectual functioning (FSIQ between 70-78; $n = 10$) or MR (FSIQ between 60-69; $n = 20$ or FSIQ between 50-59; $n = 9$) and found that the tests in this battery were failed far more often than expected (SIRS = 53.8%; TOMM = 41%; Memory for Fifteen Items Test = 79.5%; Dot Counting Test = 2.6%). It is possible, that the lower scores obtained by Hurley and Deal (2006) compared to Simon (2007) are the result of more cautious responding in individuals with mild MR who have been adjudicated. Overall, however, these data indicate that SVTs and embedded indices designed to detect malingering, primarily in TBI litigants, may not sufficiently assess feigned MR in criminal forensic settings.

Multiple Administrations of Tests and Practice Effects

In the forensic arena, criminal defendants and civil litigation plaintiffs are likely to be initially evaluated by experts hired by the defense attorney and plaintiff attorney, respectively. Criminal prosecutors and insurance companies are entitled to obtain independent medical evaluations as well. Consequently, multiple administrations of popular psychometric tests are common. With regards to the assessment of intellectual functioning, research has found that the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III), is the most commonly used instrument (Rabin, Barr, & Burton, 2005). The WAIS-III, like most commonly used assessment instruments, is susceptible to practice effects upon multiple administrations. Practice effects are the result of both memory for test-specific items and memory for procedures of the tasks (Rapport, Brines, Axelrod, & Theisen, 1997). Items most susceptible to practice effects include those that require speeded responses and those that have a single, memorable solution (Lezak, 1995). Studies have reported increases in FSIQ scores that range from 2 to 6 points during the first six months following initial evaluations (Tulsky & Zhu, 1997; Basso, Carona, Lowery, & Axelrod, 2002). These finding highlight the necessity of developing embedded validity indices for the Stanford Binet-5 designed and validated specifically for detecting feigned MR.

Rationale for Present Study

Where MR is in question and independent medical evaluations (or other repeat testing for whatever reason) are likely, the professional must mitigate practice effects whenever possible. Intelligence quotients below 70 are needed for a diagnosis of MR. Scores below 70 are obtained either because an individual actually has MR or an individual provides deviant effort. Many of the stand alone effort tests as well as the Wechsler scales have come up short in the evaluation of mentally retarded individuals. The SB5 currently has no validity indicators and establishing such would greatly increase its utility in forensic settings. A methodology is proposed that will aim to distinguish individuals feigning MR from effortful participants responding honestly and also from individuals with diagnosed MR.

Individual items from the SB5 will be examined to determine items that are rarely missed by individuals with MR and individuals exerting adequate effort. These items will be tested to identify a subset that is missed at higher frequencies by individuals attempting to feign MR. The rationale for this approach is that individuals "guessing" at how MR patients would perform are likely to answer some items in a manner that is not consistent with a majority of MR individuals. The primary goal of this study is to establish multiple items that provide high specificity so that no individual with MR is classified as malingering, while still providing sufficient sensitivity to ensure that, more probably than not, fakers still get caught.

RESEARCH QUESTIONS AND HYPOTHESES

Question 1

How will analog MR malingerers' SB5 FSIQ compared to college controls and individuals with MR?

Hypothesis 1

Analog MR malingerers will obtain FSIQs significantly lower than college controls but consistent with MR individuals from the standardization sample.

Question 2

Are there items in the standardization data that are rarely missed by individuals with MR and normal individuals?

Hypothesis 2

There are items that are rarely missed by both individuals with MR and normal individuals identified in the standardization data.

Question 3

Will analog MR malingerers fail more of the Stanford Binet-5 Rarely Missed Items (SBRMI) compared to controls and individuals with MR?

Hypothesis 3

The SBRMI will be failed more frequently by individuals asked to feign MR than controls or individuals diagnosed with MR.

Question 4

How will analog MR malingerers perform on stand-alone effort indices compared to college controls?

Hypothesis 4

Analog MR malingerers will obtain significantly lower scores on the TOMM, WMT, and Reliable Digit Span subtest compared to controls. This performance will fall below empirically derived and published cutoff scores for each measure based on the extant literature on real and simulated neurological dysfunction.

Question 5

How will the SBRMI compare with the TOMM, WMT, and Digit Span?

Hypothesis 5

The SBRMI is expected to demonstrate more sensitivity compared to the TOMM, WMT, and Digit Span and greater specificity in an MR malingering population and normal, effortful controls.

METHOD

Participants

Standardization Sample. Dr. Gale Roid, author of the SB5, provided standardization data from 343 healthy control individuals aged 17-35. Of these individuals, 14 were classified into the MR group based on intellectual deficits falling in the MR range (IQ < 70). It should be noted that a score of 70 and below was used to classify these individuals as many states use strict cutoffs in Atkins cases. Also, data from 22 individuals in the healthy control sample were deleted due to missing data. Of the 307 individuals retained for the healthy standardization control group, 51.5% were female and 48.5% were male, 70.7% were Caucasian, 12.4% were African American, 10.7% were Hispanic, 3.3% were Asian, and 2.9% were of other ethnicity. The mean age of individuals in the standardization control group was 24.11 ± 5.19 (range 17-35 years). The mean FSIQ was 102.47 ± 14.18 (range 74-146).

Dr. Roid also provided standardization data from 38 mentally retarded individuals aged 2-20. Data were retained for individual's aged 10 years and older and whose FSIQ was above 50 (n = 17). Of the individuals classified as MR by Dr. Roid, four individuals obtained FSIQ > 70 (74, 75, 76, and 95, respectively). Data from these 17 individuals were combined with data from the 14 individuals from the standardization sample who obtained scores < 71. Therefore, a total of 31 individuals with intellectual deficits were retained for this study, and 51.6% were female and 48.4% were male, 35.5% were African American, 29.0% were Hispanic, 25.8% were Caucasian, 6.5% were Asian, and 3.2% of were other ethnicity. The mean age of the MR group was 18.35 ± 6.67 (range = 10-34).

College Sample. A sample of 108 student volunteers from a large university in the southeast United States was asked to participate in this study. Individuals were undergraduates receiving extra credit in introductory psychology classes in exchange for their participation in this study. Students between the ages of 18 and 48 (mean age = 20.48 ± 4.11) with no history of a neurological disorder or current psychological disorder were eligible to participate in this study. Of this sample, 63.9% were male and 36.1% were female, 75.9% were Caucasian, 13.9% were African American, 5.6% were Asian, 1.9% were Hispanic, and 2.8% were of other ethnicity. Participants were randomly divided into two groups: analog malingerers and normal controls.

Of the individuals randomly assigned to the analog MR group, 61.1% were male, 79.6% were Caucasian, 13.0% were African American, 3.7% were Asian, 1.9% were Hispanic and 1.9% were of other ethnicity. Ages of the analog MR participants ranged from 18 to 48 years (mean: 20.96 ± 5.28), and mean years of education was 13.31 ± 1.20. The mean estimated FSIQ based on the Shipley total score was 104.74 ± 8.27 (range = 85-118).

Demographics for the college control group were as follows: 66.7% male, 72.2% Caucasian, 16.7% African American, 7.4% Asian, 1.9% Hispanic and 1.9% of other ethnicity. Ages of the analog MR participants ranged from 18 to 29 years (mean: 20.00 ± 2.411), and mean years of education was 13.13± 1.23. The mean estimated FSIQ based on the Shipley total score was 106.17 ± 7.52 (range = 87-118).

Materials

Consent Form and Demographic Questionnaire. For consent form and Demographic Questionnaire, see Appendices B and C, respectively. The consent form was the only document containing participants' names and signatures and this information was used only to ensure that they receive extra credit. It was kept in a separate folder in a secure location and no documentation was kept to associate participants' names with their test performances. All other information was kept anonymous. The demographic questionnaire obtained information such as age, education, race, sex, and any current diagnosis psychological or neuropsychological conditions.

Shipley Institute of Living Scale (Shipley). The Shipley (Shipley, 1940) is divided into a verbal and an abstraction subtest. The Verbal subtest consists of 40 words with four choices for each word. The individual must choose a synonym for the word given amongst three foils. The Abstraction subtest consists of 20 problems which the individual must solve beginning with simple problems and increasing in complexity. One point is given to each correct answer. The Abstraction raw score is multiplied by two to give the Abstraction score. Scores can be converted into estimated WAIS-R IQ scores. Participants were given 10 minutes to complete the Verbal portion and 10 minutes to complete the Abstraction portion of the test.

Stanford Binet-5. The SB5 was published in 2003 and is a broad measure of intellectual and cognitive ability (Roid, 2003a, Roid 2003b, Roid 2003c). The test has been normed for ages 2-85+ years and was standardized based on a regionally stratified national sample. The SB5 is composed of five domains measured by both verbal and nonverbal subtests. The domains are Quantitative Reasoning, Fluid Reasoning, Knowledge, Visual-Spatial Processing, and Working Memory. Individuals are first administered a verbal routing subtest and a nonverbal routing subtest to determine starting points on subsequent verbal and nonverbal subtests, respectively. In adults, each of the five domains is measured by six testlets with each testlet containing six possible points. Testlets increase in difficulty as the test progresses, and basal and ceiling rules are operationalized in the test manual using the testlet as the unit of analysis. The ceiling is determined by an individual scoring two or fewer points on a given testlet, and the basal is established when an individual scores more than two points on a testlet.

The SB5 has strong psychometric properties. Test-retest reliabilities ranged from .66-.93 for the verbal and nonverbal subtests and .89-.95 for the full test IQ scales. Correlations with the WAIS-III factor indexes range from .69-.80. Administration time typically ranges between 45-75 minutes.

Test of Memory Malingering (TOMM; Tombaugh, 1996). For the purposes of the present study, only the TOMM Trials 1 and 2 were administered (Tombaugh, 1996). During Trial 1, participants were given two practice trials before beginning testing. Following the practice trails, the participants were shown 50 line drawings at a rate of three seconds each. Individuals were then presented with pairs of line drawings; one target and one foil, and were asked to indicate which one they have previously seen. They received feedback pertaining to the accuracy of their response on each item. Trail 2 began with the presentation of the same 50 line drawings in a different order and was followed by the presentation of 50 pairs of drawings, one target and one foil. Again, participants were given feedback on the accuracy of their responses. A cutoff score of 45 (90% accuracy) on Trial 2 is recommended for the detection of poor effort.

The TOMM was normed first using 405 individuals who were part of a study on aging. The ages of participants ranged from 16-84 with a mean age of 54.8. The mean score obtained on Trial 1 was 47.5 (SD = 3.2), and the mean score for Trial 2 was 49.8 (SD = .8). A second phase in the validation process included 70 neurologically intact volunteers ranging in age from 17-73 (Mean = 37.8, SD = 14.2). Only 3% scored less than 49 on Trail 2. The test was validated in clinical patients referred for neurological assessment (n = 138) and individuals with head injuries (n = 23; Rees, 1996). All groups responded to Trail 2 with 97% accuracy except the dementia group which obtained 92% accuracy. Further validation with analog malingers (n = 27) and patients at-risk for malingering (n = 11) indicated much lower scores in malingering individuals compared to neurological patients, suggesting that the TOMM is a valid and useful test of malingering.

Green Word Memory Test (WMT). The WMT is computer administered and computer scored. Participants were presented with 20 pairs of words. During the Immediate Recognition (IR) test participants were shown 40 word pairs and asked to choose the 20 target pairs from 20 foils (foils consist of one target word and one word not previously seen). Each time, feedback was provided on the accuracy of the response. Delayed Recognition (DR) began following a 30 minute delay and consisted of the presentation of 40 pairs of words consisting of 20 target pairs and 20 foils. The IR and DR subtests are used in the detection of malingering. The computer program identifies performance as "pass", "fail", or "caution". The cutoff score of 82.5% is considered "very conservative". Scores ranging from 83-90% invoke "caution" because they are correlated with significantly lowered performance on a variety of neurological tests (Green, 2003; Green et al, 2001). Green (2003) reports sensitivity of 97.7% and a specificity of 100% in exaggerators. In order to fail the immediate recall or delayed recall effort indexes, an individual must score 3 standard deviations below the level typical of people with severe brain injuries.

Several other subtests are administered including the Multiple Choice, in which the individual is presented with the first word of a target pair and is asked to choose the matching word from 8 choices, the Paired Associates subtest in which the participant is asked to generate the second word of a learned pair from memory, and the Delayed Free Recall subtest in which the participant is asked to generate as many pairs as possible from memory. These subtests are not used in the calculation of effort bias, but are of potential value reflecting various aspects of verbal memory functions.

Digit Span Subtest of WAIS-III. Participants were asked to complete the Digit Span subtest of the WAIS-III as an additional effort measure for the computation of reliable digits. A trial consists of two strings of numbers the same length. In the Forward portion, trials are presented in increasing strings at a pace of one number per second beginning with two numbers and going up to eight numbers. Immediately after the examiner presents a string, the participant will be asked to repeat the string in the same order. The subtest is discontinued when an individual incorrectly recalls both strings in a given trial. The Digit Span Backward task is similar to the Forward task except individuals are asked to repeat the string backward, in reverse order of the examiner. The subtest is discontinued after inaccurate responses to both strings at a given trial. Reliable Digit Span scores are determined by calculating the sum of the longest digits forward and the longest digits backward. Scores below seven are indicative of poor effort.

Participant Effort-Rating Scale. Following the conclusion of testing, participants in the analog MR malingering condition were asked to complete the Participant Effort-Rating Scale (Appendix C). Participants were asked to indicate on a Likert Scale of 1 to 5 (0 = not at all; 3 = somewhat; 5 = very much so) how much effort they put forth in performing as someone with intellectual disability and how successful they thought they were at performing at that level.

Procedure

Standardization Sample. Riverside Publishing Company collected a norming sample from four geographic regions in the United States (Northeast, Midwest, South, and West). The publishers used two methods in obtaining the norming sample. First, they enlisted approximately 100 examiners from each of the predefined regions and provided training in the test's administration. Second, full-time examiners were recruited (8-10 teams per area) and tested cases of certain stratification variables in the sample (age, sex, socioeconomic status, and geographic region). The sample was matched based on percentages of stratification variables in each region based on information from the U.S. Census Bureau in 2001. The assistance of the Riverside Publishing Company and Dr. Roid is gratefully acknowledged.

College Sample. Participants were first provided with a copy of the consent form (Appendix B) which was read aloud. Afterward, questions and concerns were addressed and signatures of willing participants were collected. After informed consent was obtained, individuals were randomly assigned to the control group or the analog malingering group and were assigned an identification number to ensure anonymity. All individuals were asked to complete a demographic questionnaire, and the Shipley Institute of Living Scale giving their best effort in order to estimate each participant's FSIQ. Participants assigned to the control condition were asked to read and follow one scenario while participants assigned to the analog malingering condition were asked to read and follow a different scenario. The scenarios were balanced for reading skill and word count, but differed in information and instructions for approaching the remaining tests. After instructions were explained, all subjects were administered the WMT (IR), TOMM Trials 1 and 2, Digit Span, WMT (DR) and the SB5 in a counterbalanced order. Administration of all items was performed by trained examiners. Administration time was approximately 3 hours per individual. After completing the evaluation, participants were provided with information about the purposes and utility of the study. Individuals in the analog MR malingering condition were asked to complete the Participant Effort Rating Scale. Each SB5 protocol was scored independently by two examiners (Mandi Muss and Alyse Barker), and inconsistencies were evaluated.

Scenarios.

Control Condition. Individuals who were assigned to the control group were asked to read the following passage:

> "As a child, you were able to complete schoolwork, read, write, spell, and do math as well as other children. In addition, you were able to adequately perform daily functions related to interpersonal relationships, monetary transactions, communication, and self-care.
>
> As you grew older, you perceived yourself as "normal" and were never diagnosed with an intellectual disability. You graduated from high school and were accepted to Louisiana State University. You began hanging around people who convinced you that obtaining a college education was the best way to support yourself. One night you decided to sign up for a psychology course at LSU. The psychology course offered you an opportunity for research credit or extra credit by participating in psychology experiments. You chose to sign up for this psychology experiment.
>
> Now you are sitting in the LSU Psychological Services Center participating in this study for research credit or extra credit. You are asked to take some tests as part of a neuropsychological evaluation. The examiner explains that, if you perform with your best effort, you will contribute to psychological research. The examiner further explains that psychological experiments are crucial to testing hypotheses and developing theories.
>
> The validity of this research study depends on the effort you put forth on these tests. You are instructed to respond to the tests with your very best effort" (A. A. Barker, personal communication, October 30, 2009).

Analog Malingerers. Individuals assigned to the analog malingerer group were asked to read the following brief scenario:

> “As a child, you struggled continuously with schoolwork and were never able to read, write, spell, or do math as well as other children. In addition, you struggled with daily functions related to interpersonal relationships, monetary transactions, communication, and self-care.
>
> As you grew older, you knew ‘something was wrong.’ You either ‘squeaked by’ though school, got socially promoted, or dropped out. You continued to struggle with basic skills. You began hanging around people who convinced you that stealing was the best way to support yourself. One night you broke into a house to steal some money, and the homeowner confronted you with a gun. In a split second reaction, you shot him fatally.
>
> Now you are in prison, facing the death penalty. You are asked to take some tests as part of a neuropsychological evaluation. Your lawyer explains that, if the tests show your IQ is less than 70, you can be found mentally retarded and escape the death penalty, as mentally retarded persons cannot be executed. Your lawyer further explains, however, that, if the examiner discovers you are faking, you will likely receive the death penalty.
>
> Your life is now depending on the scores of these tests. You are instructed to respond to the tests in a manner that ensures you will be found mentally retarded without being detected as a faker” (A. A. Barker, personal communication, October 30, 2009).

Note that participants were explicitly asked to answer the following materials by feigning mild MR in a way that was not obvious to the examiner. While studies utilizing analog malingerers in TBI typically have shorter scenarios (Swihart et al., 2008; Mittenberg et al., 1995), it was essential for participants in the current study to understand lifelong complications involved in mild MR and to understand the degree to which individuals with mild MR are able to function in society. The control scenario was balanced for length with the analog scenario.

Statistical Analyses

A power analysis utilizing G*Power 3.1.0 (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007) revealed that a Chi-square test used to detect a medium effect size (w =.35) in a total sample of 130 when alpha = 0.05, yields a power of .88. SPSS 18.0 was used in all comparisons between and within groups. Chi-square tests were used to compare race and Fisher's Exact tests were used to compare gender across groups. T-tests were used to compare age, years of education, and estimated WAIS-III FSIQ scores between the college controls and the college analog participants. Mann-Whitney U tests were used to compare age and SB5 FSIQ scores between analog participants and MR individuals. Logistic regression was used to investigate the predictive power of the SBRMI and ROC curves were used to examine SBRMI characteristics. In addition, χ^2 analyses were used to further examine utility of the SBRMI in differentiating analog malingers from individual with MR, college controls, and the standardization sample. Equations for percentages of items missed per subject, sensitivity, specificity, Positive Predictive Power (PPP), and Negative Predictive Power (NPP) were calculated by hand using the equations provided in Baldessarini, Finklestein, and Arana (1983; Table 2). Pearson's *r* correlations and Mann-Whitney U tests were used to examine the relationship between the effort rating scale and performance on measures of malingering. Also, logistic regression and diagnostic statistics were examined for the TOMM, RDS, and WMT IR and DR subtests to compare college controls to analog participants.

Table 2. Equations for predictive analyses of the SBRMI.

SBRMI Classification	Analog MR	MR	Equations
Positive for Malingering	a	b	**PPP = a/(a + b)**
Negative for Malingering	c	d	**NPP = d/(d + c)**
	Sensitivity = a/(a + c)	**Specificity = d/(b + d)**	

RESULTS

Demographic Information

A Pearson Chi-square test indicated that there was no significant difference among analog MR and college student control groups for race/ethnicity ($\chi^2 (1) = 1.11$). A Fisher's Exact test found no significant gender differences (p (1) = .689) between analog MR subjects and college controls. Independent samples t-tests indicated no significant differences in age (t (106) = 1.22; $p < .226$), years of education (t (106) = .789, $p < .432$) or estimated WAIS-R IQ (t (106) = 1.25; $p < .21$) scores. When analog MR malingerers were compared to MR individuals from the standardization sample, there was a significant difference in race ($\chi^2 (1) = 22.74$; $p < .001$). There were significantly fewer Caucasians and significantly more African Americans and Hispanics in the actual MR sample. A Fisher's Exact test revealed no significant differences in gender ($p = 0.34$) for the analog MR compared to actual MR groups. Mann-Whitney U tests revealed that there was a significant difference in the age ($z = 2.32$; $p < .05$) of the analog MR group compared to the actual MR group with the analog MR individuals being older.

Hypothesis 1

The mean analog MR SB5 FSIQ was 66.13 ± 20.56 (range = 40-110) while the mean SB5 FSIQ score for college controls was 104.48 ± 10.91 (range = 78-126). A Mann-Whitney U test revealed that analog MR participants obtained significantly lower FSIQ scores on the SB5 compared to college controls ($z = 7.90$; $p < .001$). Mean FSIQ scores for MR participants of the SB5 standardization data ranged from 52 to 95 (mean = 63.68 ± 8.49) and were not significantly different compared to analog participants ($z = .18$; $p < .855$).

Hypothesis 2

The pool of potential SBRMI items was obtained by calculating the least frequently missed items in the SB5 MR standardization data. The SBRMI consisted of items missed by analog malingers significantly more ($p < .001$) often than MR individuals. Twenty-one non-verbal items with $p < .001$ were retained for the SBRMI. Each item was summed to obtain the total SBRMI score (Item and Chi-square statistics found in Table 3). Essentially, a score of 21 indicates that an individual responded to all SBRMI items correctly and a score of 0 indicates that an individual missed every item of the SBRMI.

Table 3. Rarely missed items of the SB5

SB5 Items	ROID MR (n = 31)		LSU AMR (n = 54)		Pearson's χ^2	Fisher's Exact Two-sided
	0pt	1pt	0pt	1pt		
NV Routing-7	2	29	21	33	10.50	0.001
NV Routing-8	0	31	16	38	11.32	0.001
NV Routing-12	0	31	16	38	11.32	0.001
NV Routing-13	0	31	22	32	17.04	0.001
NV Routing-14	0	31	20	34	15.01	0.001
NV Routing-15	2	29	26	28	15.50	0.001
NV Routing-16	0	31	20	34	15.01	0.001
NV Routing-17	1	30	27	27	19.51	0.001
NV-QR Level 2-4	0	31	17	37	12.20	0.001
NV-VS Level 2-3	0	31	14	40	9.62	0.001
NV-VS Level 2-5	0	31	14	40	9.62	0.001
NV-VS Level 2-6	0	31	17	37	12.20	0.001
NV-WM Level 2-4	0	31	14	40	9.62	0.001
NV-WM Level 2-5	0	31	18	36	13.11	0.001
NV-WM Level 2-6	0	31	21	33	16.01	0.001
NV-QR Level 3-3	0	31	26	28	21.50	0.001
NV-QR Level 3-4	0	31	24	30	19.20	0.001
NV-WM Level 3-1	1	30	27	27	19.51	0.001
NV-WM Level 3-2	0	31	27	27	22.72	0.001
NV-WM Level 3-3	3	28	30	24	17.45	0.001
NV-KN Level 4-1	8	23	35	19	11.99	0.001

Hypothesis 3

A logistic regression analysis indicated that the SBRMI was a significant predictor of malingered mild MR (n = 85; χ^2 (1) = 41.18; p < .001), Odds Ratio = 1.92 CIs [1.27, 2.98]. Overall, classification was adequate with 77.6% accuracy. Specifically, 74.1% of analog malingers identified as such and 83.9% of actual MR individuals were identified correctly when the SBRMI total score was entered into a logistic regression as the predictor. When individuals who failed to suppress their FSIQ scores below 85 were excluded, the SBRMI remained a significant predictor of malingering (n = 75; χ^2 (1) = 56.87; p < .001), Odds Ratio = 2.69 CIs [1.56, 4.63] and overall classification improved for the analog group (88.6% correctly classified).

The author used an ROC curve to examine the tradeoff of sensitivity and specificity of the SBRMI for clinical purposes. Ultimately, the author was seeking a cutoff score that was highly specific in detecting individuals with mental retardation. Results of the ROC indicated that a cutoff score of 17.50 (averaged between scores 17 and 18) yielded sensitivity of 59.3%, a specificity of 100%, a PPP of 100% and a NPP of 58.5% when all 54 analog malingerers were used despite SB5 FSIQ. When only individuals who suppressed their SB5 FSIQ scores below 85 were used (n = 44; Figure 1), a cutoff score of 17.50 yielded 70.5% sensitivity, 100% specificity, 100% PPP and 70.5% NPP .

Chi-square tests were used to investigate the utility of a cutoff score of 17 on the SBRMI. When all analog participants were included, a cutoff score of 17 correctly identified 32 of 54 analog malingerers and all 31 mentally retarded individuals. When individuals who failed to suppress their FSIQ scores on the SB5 were excluded (n = 10), a cutoff score of 17 correctly identified 31 of 44 analog malingerers. Fisher's exact tests indicated that the analog malingerers (n = 54; χ^2 (1) = 37.23; p < .001 and n = 44; χ^2 (1) = 29.46; p < .001) differed significantly from actual MR individuals when a cutoff score of 17 was used.

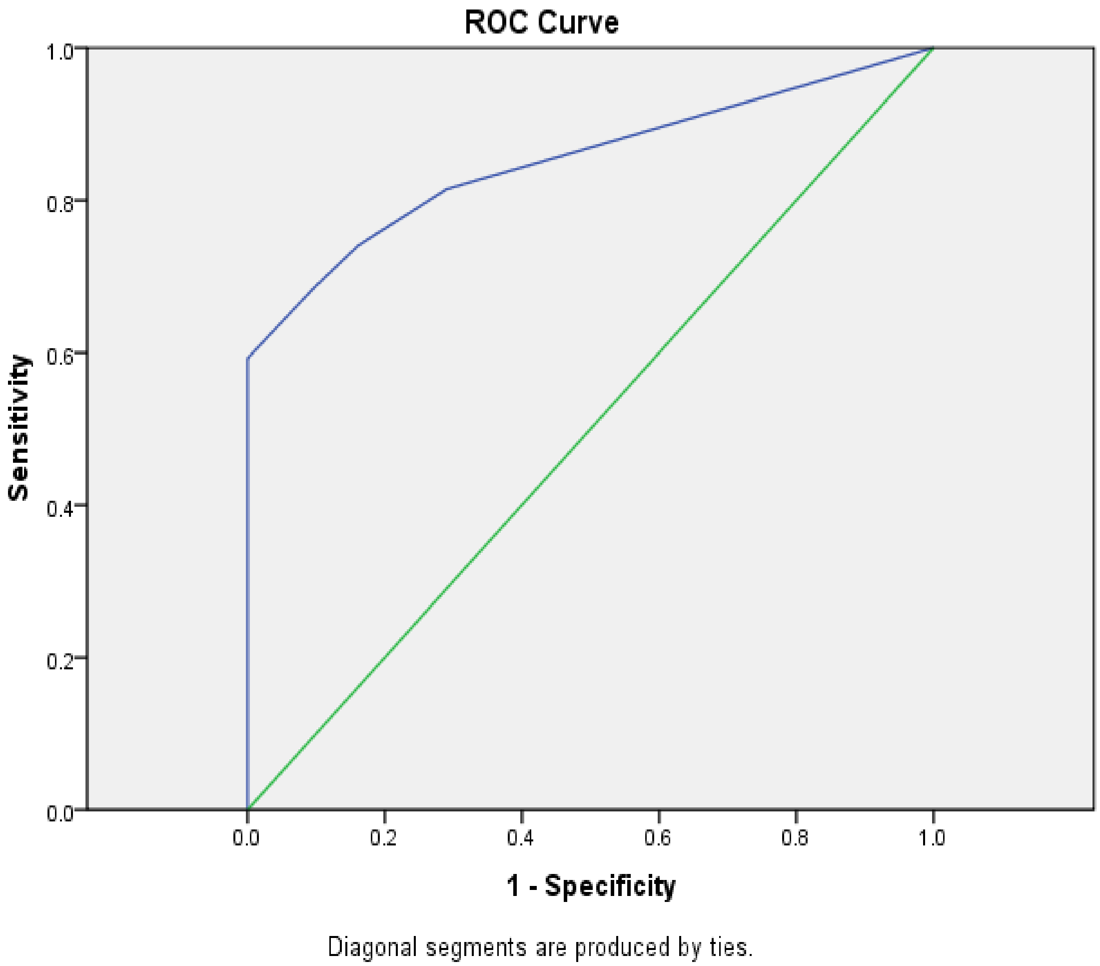

Figure 1. ROC curve analysis for SBRMI indicating good sensitivity and excellent specificity to malingered mild mental retardation in a sample of MR individuals and college simulators.

Hypothesis 4

Mann-Whitney U tests were used to compare differences in scores on the TOMM, WMT, BSRI, and Digit Span between individuals in the analog MR group and college controls (means and standard deviations presented in Table 4). The analog MR group's scores on the TOMM Trail 1 (z = 7.28; p < .001), TOMM Trial 2 (z = 6.96; p < .001), Reliable Digits Index (z = 5.41; p < .001), WMT Immediate Recall (z = 7.32; p < .001) and

WMT Delayed Recall ($z = 7.83$; $p < .001$) tasks were significantly lower compared to college controls. The BSRI was used as a measure of discriminant validity; however, analog MR participants obtained significantly different scores on the BSRI compared to college controls indicating that they chose to respond differently on this measure of masculinity/femininity: BSRI Masculine ($z = 2.50$; $p < .01$) and Feminine ($z = 2.61$; $p < .01$). It should be noted, that analog participants' responses were more than two and a half standard deviations higher on the Masculine and Feminine scales of the BSRI compared to controls.

Table 4. Mean ± SD scores for analog participants and college controls.

	LSU Analog MR Mean ± SD	**LSU Controls Mean ± SD**	**Significance**
Shipley IQ	104.74 ± 8.27	106.17 ± 7.52	.351
SB5 FSIQ	66.37 ± 20.55	104.56 ± 10.82	.001
TOMM T1	34.19 ± 10.73	48.48 ± 2.34	.001
TOMM T2	37.26 ± 10.74	49.918 ± 0.35	.001
RDS	6.91 ± 3.23	10.31 ± 3.11	.001
WMT IR	70.63 ± 19.46	96.53 ± 12.03	.001
WMT DR	70.48 ± 19.27	98.58 ± 3.11	.001
BSRI-M	4.39 ± 1.30	4.96 ± 0.64	.01
BSRI-F	4.26 ± 0.87	4.71 ± 0.61	.01

Hypothesis 5

The WMT IR and DR, TOMM Trail 2, SBRMI, and RDS scores were entered into a logistic regression analysis. Overall, these symptom validity tests were significant predictors of malingered MR ($n = 108$; $\chi^2 (5) = 109.35$; $p < .001$). However, further examination of indicated that only the WMT DR ($p < .01$; Odds ratio = 1.27) and SBRMI ($p < .01$; Odds ratio = 8.83) significantly predicted malingered MR. The WMT IR ($p < .815$; Odds ratio =

1.01), TOMM Trial 2 ($p < .07$; Odds ratio = 1.59), or the RDS ($p < .37$; Odds ratio = .81) were significant predictors. Overall, classification was excellent with 93.4% accuracy. Specifically, 90.7% of analog malingers identified as such and 96.2% of actual MR individuals were identified correctly when the WMT IR and DR, TOMM Trail 2, and RDS scores were entered into a logistic regression as predictors.

Chi-square analyses revealed that the SBRMI cutoff score of 17 correctly identified all of the college controls ($\chi^2(1) = 45.47$; $p < .001$) and all of the healthy individuals from the standardization sample ($\chi^2(1) = 199.62$; $p < .001$). Sensitivities and specificities were calculated in order to determine whether the TOMM, WMT, SBRMI, and RDS were able to differentiate college controls from analog malingerers. A cutoff score of 44 on the TOMM Trial 2 demonstrated 100% specificity and 63% sensitivity in differentiating analog participants from college controls. A cutoff score of 81.50 for the WMT IR index demonstrated 98.1% specificity and 68.5% sensitivity, and cutoff score of 79 demonstrated 64.8% sensitivity and 100% specificity was obtained. A cutoff score of 17.5 on the SBRMI yielded 59.3% sensitivity and 100% specificity when all 54 analog malingerers were included. A reliable-digits cutoff of 6.5 yielded 96.3% specificity and 50% sensitivity. When only analog participants with FSIQ score below 85 were used, sensitivities were 72.7%, 61.4%, 72.7%, 72.7%, and 70.5% for the TOMM Trial 2, RDS, WMT IR, WMT DR, and SBRMI and specificities were 100%, 96.3%, 98.1%, 100%, and 100%, respectively.

Additional Analyses

As an additional analysis, not originally proposed for this thesis, validity of the effort rating scale was examined. Pearson's r correlations indicate a significant relationship between reported effort on the effort rating scale and SB5 FSIQ (r = -.35; p < .01), SBRMI (r = -.32, p < .05), TOMM Trail1 (r = -.30; p < .05), TOMM Trial 2 (r = -.37; p < .01). There were no significant relationships between the effort rating scale and the RDS (r = -.18; p < .21), WMT IR (r = -.16; p < .26), WMT DR (r = -.20; p < .15), BSRI masculine (r = -.03; p < .83), or BSRI feminine (r = -.01; p < .95) scores. The analog MR group was divided into individuals who reported giving average or below average effort and individuals who reported giving above average effort. Mann Whitney U tests indicated that there were no significant differences in scores on the TOMM Trail 1 (z = 1.14; p < .254), TOMM Trial 2 (z = 1.40; p < .162), Reliable Digits Index (z = .267; p < .789), and Word Memory Test Immediate Recall (z = 1.51; p < .291), and Delayed Recall (z = .647; p < .518), tasks in the groups who reported below average/average effort compared to those who reported above average effort. Individuals in both groups also obtained similar scores on the BSRI: Masculine (z = .751; p < .453), Feminine (z = .532; p < .594), and Neutral (z = .323; p < .747). Scores on the SB5 FSIQ did not differ significantly (z = 1.75; p < .080) between individuals who reported giving above average effort and those who reported average/below average effort. However, students who reported giving above average effort during the study scored significantly lower on the SBRMI compared to individuals who reported giving average to below average effort (z = 2.85; p < .01) based on reported effort measured by the effort rating scale. The mean SRRMI score for individuals who reported giving above average effort was 10.85 ± 7.225 whereas the mean score for participants who reported giving average/below average effort was 16.37 ± 6.48.

DISCUSSION

The present study endeavored to derive a rarely missed items index from the Stanford Binet-Fifth Edition (SBRMI) designed specifically to detect malingered mental retardation. This study employed college students asked to simulate mental retardation and college controls. In addition, a sample of healthy individuals and a sample of mentally retarded individuals from the SB5 standardization data were used. An item analysis revealed 21 items by which analog MR participants differed significantly ($p < .001$) from actual MR individuals. These items were summed into a total score, the SBRMI. Logistic regression analyses revealed that the SBRMI was a significant predictor of malingered mental retardation and ROC analyses revealed adequate sensitivity and excellent specificity when a cutoff score of 17 was used. Thus, hypotheses 2 and 3 were confirmed.

Killgore and DellaPietra (2000) used discriminate function analyses with all six items from the WMS-III RMI index as predictors of group membership (i.e. simulated head injury or control). However, because scores in the present study were not normally distributed logistic regression analysis was used to determine the predictive utility of the SBRMI. Logistic regression functions in a similar manner to discriminate function analysis; however, it does not have assumptions regarding normality or equal variance, and it allows for binomial responses (Tabachnick & Fidell, 2007). In addition to the logistic regression analyses, three pre-planned χ^2 comparisons (analog MR and standardization MR, analog MR and college controls, and analog MR and healthy standardization sample) of the SBRMI total scores were examined.

Analog malingerers varied widely in response styles adopted for the SB5. However, hypothesis 1 was confirmed as the mean SB5 FSIQ did not differ significantly from FSIQ scores obtained by mildly mentally retarded individuals but was significantly lower compared to the college controls. Ten analog participants obtained FSIQ scores above 85 indicating that they failed to adequately follow the instructions. Information from these subjects was retained in initial analyses, however, in order to provide more robust data.

It is noteworthy that when the ten individuals with FSIQ > 85 were removed from analyses, classification rates of the SBRMI increased substantially. In clinical practice, individuals assessed for mental retardation are likely to have actual FSIQ scores that fall in the borderline intellectual functioning range or mild MR range. It is relatively unlikely that an individual with a FSIQ score in the average range would be referred for evaluation of mental retardation. Therefore, it is promising that classification increased when individuals with higher IQ scores were removed. It is also important to note that no MR individuals were incorrectly classified as malingering when a cutoff score of 17 was applied. Individuals with MR ranged in age from 10-34, yet even younger children with MR did not fail the SBRMI. In addition, none of the college controls or healthy individuals from the standardization sample (IQ scores ranged from 74-146) were classified incorrectly as malingering mild MR even though some individuals from those samples obtained IQ scores that fell in the borderline to low average ranges.

Findings that analog malingerers differed significantly from college controls on the TOMM, WMT, and RDS are consistent with hypothesis 4. These analyses were included as a manipulation check and suggest that analog malingerers chose to distort their responses on standalone effort measures. Surprisingly, analog malingerers also differed from college controls on the BSRI, a measure of masculinity and femininity. The scores of analog malingerers were lower for both the masculinity and femininity scales. It is unclear why analog malingerers would chose to responded differently on a measure of sex roles, but it is possible that the analog malingerers chose a more careless response style or chose not to respond as honestly as control subjects. An alternative hypothesis is that the analog malingerers chose to respond as poor readers might and these significant differences reflect effort to feign MR on the BSRI.

Hypothesis 5 was not confirmed as the TOMM Trial 2 and WMT IR and DR indexes yielded slightly better sensitivities and comparable specificities when comparing malingered MR and college controls. Unfortunately, the sensitivity and specificity of the SBRMI could not be directly compared to the TOMM, WMT, and RDS in the MR sample because symptom validity test scores were not available for the data provided by Dr. Roid. Items of the SBRMI were chosen based on infrequency of incorrect responses in children and adults who obtained intellectual deficits that placed them in the mild MR range. These criteria for the MR group add to the validity of this index because even children and adolescents with mild MR missed the selected items infrequently. Therefore, it is hypothesized that the SBRMI, designed specifically to detect feigned MR, would demonstrate better sensitivity and specificity compared to other effort measures as the other measures used were designed to detect cognitive impairment, most frequently, feigned traumatic brain injury. Studies have reported conflicting results about the utility of the TOMM, RDS, and WMT in samples of individuals diagnosed with mental retardation (Simon, 2007; Graue et al, 2007, Marshall & Happe, 2007), and overall, examination of the literature indicates that current symptom validity tests and embedded indices are not valid in an intellectually disabled population, or at the very least, new cutoff scores need to be derived.

In additional analyses, this study attempted to measure the amount of effort that students put forth in attempting to feign mild mental retardation by incorporating an effort rating scale. The scale consisted of two questions evaluating how hard the individual tried and how well they thought they did. Analysis of the effect of effort as measured by the effort rating scale indicates that self-reported effort was negatively correlated with test scores. Mann-Whitney U tests indicated no significant differences on any of the symptom validity tests or on the SB5 FSIQ scores. However, individuals who reported giving above average effort scored significantly lower on the SBRMI compared to those who reported giving average/below average effort. It is unclear how student's conceptualized the question: how hard did you try. It appears that they may have based their answers on how many times they purposefully responded incorrectly on the SB5 as they missed significantly more rarely missed items compared to individuals with average/below average effort. Only 52 students completed the effort rating scale, and the effort rating scale needs further validation in order to be useful for these types of studies.

This study has several limitations. First, the method by which students choose to feign mental retardation has not been investigated in the literature. Therefore, the utility of the SBRMI needs further validation in MR individuals as well as settings where feigned MR is probable. Also, the sample size of MR individuals from the standardization data was small and was supplemented by individuals from the standardization data deemed healthy, but with intellectual deficits that fell in the MR range. This sample ranged in age from 10 to 34 years old with a significant percentage of individuals under the age of 18. Future research should focus on obtaining larger samples of adults with mental retardation diagnoses. Also, the SBRMI is composed only of items from the nonverbal subtests. It is unclear whether this pattern will stand when further validated in MR and probable feigned MR samples. Most importantly, as stated previously, this index is most likely going to be used on individuals with borderline intellectual functioning to low average functioning who are trying to suppress their IQ scores to the MR range. College students have IQ scores that are, on average, over one standard deviation higher than the predicted population for which this index will be used. Therefore, it is unclear whether individuals with lower IQ scores would chose to feign the same way or on the same items. However, it is hypothesized, that individuals with lower intellectual functioning who have decided to feign, would exaggerate deficits in a manner consistent with, or at least similar to college analog participants.

In conclusion, this initial study indicates that a rarely missed item index from the SB5 shows promise in differentiating mentally retarded individuals from persons attempting to feign mental retardation. The SBRMI is composed of 21 items with a total of 21 possible points. A cutoff score of 17 (i.e. $\geq$ 4 items missed) demonstrated the best sensitivity, specificity, and overall classification rates. These rates were comparable to scores obtained on symptom validity tests and the RDS, an embedded index from the WAIS-III when compared to college controls. At this point, generalizability of this index is limited and future research is needed to validate it. Future research should also expand data collection to a larger sample of adult with MR diagnoses and to setting in which feigned MR is likely. This is the first embedded index designed specifically for detection of feigned MR and validation of this index will be useful in Atkins cases as well as disability evaluations. It demonstrates excellent specificity and good sensitivity in this population. This index will expand the tool kit of clinicians and enable them to confidently assess malingering in a population of mentally retarded individuals.

REFERENCES

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.

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

Atkins v. Virginia. (2002), 260 Va. 375, 534S. E. 2d 312.

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),* 5690573.

Barker, A. A. (2009). Personal Communication.

Basso, M. R., Carona, F. D., Lowery, N., & Axelrod, B. N. (2002). Practice effects on the WAIS-III across 3- and 6-month intervals. *The Clinical Neuropsychologist. 16(1),* 57-63.

Beaber, R. J., Marston, A., Michelli, J., & Mills, M. J. (1985). A brief test for measuring malingering in schizophrenic individuals. *American Journal of Psychiatry, 142(12),* 1478-1481.

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: a dose-response relationship? Clinical Neuropsychologist, 20, 831-847.

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

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment, 3,* 129-136.

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

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegren, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring.* Minneapolis: University of Minnesota.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56(2),* 81-105.

Conder, R., Allen, L., & Cox, D. (1992). *Manual for the computerized assessment of response bias.* Durham, NC: CogShell.

D'Amato, C. P. & Denney, R. L. (2008). The diagnostic utility of the rarely missed index of the Wechsler Memory Scale-Third Edition in detecting response bias in an adult male incarcerated setting. Archives of Clinical Neuropsychology, 23(5), 553-561.

Dean, A. C., Victor, T. L., Boone, K. B., & Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 22,* 702-722.

Everington, C. & Olley, J. G. (2008.) Implication of Atkins v. Virginia: Issues in Defining and Diagnosing Mental Retardation. *Journal of Forensic Psychology Practice, 8(1), 1*-23.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

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

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: the power of external incentives. *Brain Injury, 21(4),* 373-383.

Frederick, R. I. (2000). A personal floor effect strategy to evaluate the validity of performance on memory tests. *Journal of clinical and Experimental Neuropsychology, 22(6),* 720-730.

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

Gouvier, W. D., Hayes, J. S., & Smiroldo, B. B. (1998). The significance of base rates, test sensitivity, test specificity, and subject's knowledge of symptoms in assessing TBI sequelae and malingering. In C. R. Reynolds (Ed.). *Detection of malingering during head injury litigation* (pp. 55-80). New York; 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. Prigatano & N. Pliskin (Eds.) *Clinical Neuropsychology and Cost Outcomes Research: A beginning*. New York: The Psychology Press.

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

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

Green, P. (2003). *Green's Word Memory Test for Windows user's manual.* Edmonton, Canada: Green's Publishing Company.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive Impairment* (pp. 50-77). New York: Guilford Press.

Green, P., Allen, L. M., & Astner, K. (1996). The Word Memory Test: A user's guide to the oral and computer-administered forms, US version 1.1 (Durham, NC: CogniSyst, Inc.).

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23(9),* 741-750.

Green, P., Iverson, G. L., & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13(10),* 813-819.

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

Greiffenstein, M. F. & Baker, W. J. (2006). Miller was (mostly) right: Head injury severity inversely related to simulation. *Legal and Criminological Psychology, 11,* 131-145.

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

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

Greve, K. W., Binder, L.M., & Bianchini, K. J. (2009). Rates of below-chance performance in forced-choice symptom validity tests. *The Clinical Neuropsychologist, 23,* 534-544.

Greve, K. W., Bianchini, K. J., & Doane, B. M. (2006). Classification accuracy of the Test of Memory Malingering in traumatic brain injury: results of a known-group analysis. *Journal of Clinical and Experimental Neuropsychology, 28,* 1176-1190.

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,* 245-260.

Hall, H. V. & Pritchard, D. A. (1996). Malingered neuropsychological deficits. In *Detecting malingering and deception: Forensic distortion analysis (FDA*) (pp. 69-89). Boca-Raton, FL: St. Lucie Press.

Harber, A. H. & Fichtenberg, N. L. (2006). Replication of the Test of Memory Malingering (TOMM) in a traumatic brain injury and head trauma sample. *The Clincial Neuropsychologist, 20,* 524-534.

Hayes, J. S. Hale, D. B. & Gouvier, W. D. (1997). Do tests predict malingering in defendants with mental retardation? *The Journal of Psychology, 131(5),* 575-576.

Hayes, J. S. Hale, D. B. & Gouvier, W. D. (1998). Malingering detection in a mentally retarded forensic population. *Applied Neuropsychology, 5,* 33-36.

Heaton, R. K., Smith, Jr., H. H.Lehman, R. A. W., & Vogt, A. T. (1978). Prospects of 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.

Hilsabeck, R. C., Gouvier, W. D., & Bolter, J. F. (1998). Reconstructive memory bias in recall of neuropsychological symptomology. *Journal of Clinical and Experimental Neuropsychology, 20(3),* 328-338.

Hiscock, M., & Hiscock, C. K. (1989). Refining the forced choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology, 27,* 610-617.

Hurley, K. E. & Deal, W. P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation, 44*, 112-119.

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

Killgore, W. D. S. & DellaPietra, L. (2000). Using the WMS-III to detect malingering: empirical validation of the rarely missed index (RMI). *Journal of Clinical and Experimental Neuropsychology, 22(6),* 761-771.

Lachenbruch, P. A. & Clements, P. J. (1991). Anova, kruskal-wallis, normal scores and unequal variance. *Communications in Statistics-Theory and Methods, 20(1),* 107-126.

Lange, R. T., Sullivan, K. & Anderson, D. (2005). Ecological validity of the WMS-III rarely missed index in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology, 27,* 412-434.

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

Larrabee, G. J. (2007). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios [Electronic Version]. *The Clinical Neuropsychologist,* (1 – 14). Retrieved June 19, 2008, from http://www.psypress.com/tcn

Larrabee, G. J., & Berry, D. T. R. (2007). Diagnostic classification statistics and diagnostic validity of malingering assessment. In G. J. Larrabee (Ed.). *Assessment of Malingered Neuropsychological Deficits.* (pp. 14-26). New York: Oxford University Press.

Lees-Haley, P. R. & Brown, R. S. (1993). Neuropsychological complaint base rates of 170 personal injury claimants. *Archives of Clinical Neuropsychology, 8,* 203-209.

Lees-Haley, P. R., Williams, C. W., Zasler, N. D., Marguilies, S., English, L. T., & Stevens, K. B. (1997). Response bias in plaintiffs' histories. *Brain Injury, 11(11),* 791-799.

Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

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

Meehl, P. (1956). Clinical versus actuarial predication. In *Proceedings of the 1955 Invitational Conference on Testing Problems.* Princeton: Educational Testing Service, 136-144.

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

Miller, H. A. (2001). *MFAST: Miller Forensic Assessment of Symptoms Test professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Miller, L. J.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.

Millis, S. R. & Volinsky, C. T. (2001). Assessment of response bias in mild head injury: beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology, 23(6),* 809-828.

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

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

O'Bryant, S. E. & Lucas, J. A. (2006). Estimating the predictive value of the Test of Memory Malingering: An illustrative examplefor clinicians. *The Clinical Neuropsychologist, 20,* 533-540.

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

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.

Rapport, L. J., Brines, D. B., Axelrod, B. N., & Theisen, M. E. (1997). Full scale IQ as mediator of practice effects: the rich get richer. *The Clinical Neuropsychologist, 11(4),* 375-380.

Rees, L. M., Tonbaugh, T. N., Gansler, D. A., & Moczynski, N. P. (1998). Five validation experiments of the Test of Memroy Malinering (TOMM). *Psychological Assessment, 10(1),* 10-20.

Rey, A. (1941). L'examen pscyholgique dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28,* 268-340.

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

Rogers, R. (1990a). Models of Feigned Mental Illness. *Progression Psychology: Research and Practice, 21(3),* 182-188.

Rogers, R. (1990b). Development of a new classificatory model of malingering. *Bulletin of American Psychiatry and Law, 18(3),* 323-333.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *The SIRS test manual*. Tampa, FL: Psychological Assessment Resources.

Roid, G. H. (2003a). *Stanford-Binet Intelligence Scales: Fifth Edition*. Itasca, IL: Riverside.

Roid, G. H. (2003b). *Stanford-Binet Intelligence Scales: Fifth Edition, Examiner's Manual.* Itasca, IL: Riverside.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, technical manual.* Itasca, IL: Riverside Publishing.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *The Jouranl of Psychology, 9,* 371-377.

Simon, M. (2007). Performance of mentally retarded forensic patients on the Test of Memory Malingering. *Journal of clinical Psychology, 63(4),* 339-344.

Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced choice memory test. *Journal of Clinical and Experimental Neuropsychology, 16,* 472-481.

Slick, D., Hopp, G., Strauss, E., & Spellacy, R., (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,* 911-922.

Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test.* Odessa, Fla.: Psychological Assessment Resources.

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.

Swihart, A. A., Harris, K. M., & Hatcher, L. L. (2008). Inability of the Rarely Missed Index to identify simulated malingering under more realistic assessment conditions. *Journal of Clinical and Experimental Neuropsychology, 30(1),* 120-126.

Thompson, L. L. & Cullum, C. M. (1991). Pattern of performance on neuropsychological tests in relation to effort in mild head injury patients. *Archives of Clinical Neuropsychology, 6,* 231.

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

Tombaugh, T. (2002). The Test of Memory Malingering (TOMM) in forensic psychology. *Journal of Forensic Neuropsychology*, 69-96.

Tulsky, D., & Zhu, J. (1997). *WAIS-III/WMS-III Technical Manual.* San Antonio: Psychological Corporation.

Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A Validation of the Test of Memory Malingering in a Forensic Psychiatric Setting. *Journal of Clinical and Experimental Neuropsychology, 25(7),* 979-990.

APPENDIX A: INDEX OF ABBREVIATIONS

APA	American Psychological Association
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders-4th Edition, Revised
FCT	Forced Choice Test
FSIQ	Full Scale Intelligence Quotient
IQ	Intelligence Quotient
MMPI-2	Minnesota Multiphasic Personality Inventory, 2nd Edition
MND	Malingered Neurocognitive Dysfunction
MR	Mental Retardation
NPP	Negative Predictive Power
RDS	Reliable Digit Span
PPP	Positive Predictive Power
SB5	Stanford Binet-5th Edition
SBRMI	Stanford Binet-5 Rarely Missed Items
SVT	Symptom Validity Test
TOMM	Test of Memory Malingering
WAIS-III	Wechsler Adult Intelligence Scale- 3rd Edition
WAIS-R	Wechsler Adult Intelligence Scale-Revised
WMS-III	Wechsler Memory Scale-3rd Edition
WMT	Word Memory Test

APPENDIX B: CONSENT FORM

Louisiana State University
Psychology Study Consent Form

Study Title: Simulated sub-normal performance on the Stanford Binet-: An exploratory investigation.

Performance Site: Louisiana State University Psychological Services Center

Contact Information:

Wm. Drew Gouvier, Ph.D.	Mandi Musso, B. S.	Alyse Barker, B.S.
wgouvie@tigers.lsu.edu	mmusso4@tigers.lsu.edu	abarke1@tigers.lsu.edu

This study will be looking at whether test scores can distinguish individuals' faking mild mental retardation from individuals giving their best effort. Today, you will be asked to complete a demographic questionnaire as well as several activities, including providing word definitions, identifying patterns, solving number problems, providing answers to questions, repeating information, tapping blocks, and identifying previously seen objects. These tasks will take approximately three hours to complete.

Individuals over the age of 18 who have no pending criminal charges, no significant neurological disorders, and no current psychological disorders are eligible to participate in this study. Participation is voluntary. You may at any time withdraw from the study without penalty. LSU students who participate in this study will receive extra credit in their undergraduate psychology course. There are no foreseeable risks in participating in this study; in addition, standards of psychological practice in forensic and general clinical settings may benefit from the development of indices that can be used to identify people who are faking impairment in order to obtain financial compensation or other secondary gain.

All data collected will be anonymous. Your name and identifying information will in no way be linked to your test scores.

The study has been discussed with me and all my questions have been answered. I may direct additional questions regarding study specifics to the investigators. If I have questions about subjects' rights or other concerns, I can contact Robert C. Mathews, Chairman, LSU Institutional Review Board, (225)578-8692, irb@lsu.edu, www.lsu.edu/irb. I agree to participate in the study described above and acknowledge the researchers' obligation to provide me with a copy of this consent form if signed by me.

Subject Signature:_____________________________ Date:________________
Witness Signature: ____________________________ Date: _________________

APPENDIX C: DEMOGRAPHIC QUESTIONNAIRE

Demographic Questionnaire

***Name:**
____

***Class for which you are seeking extra credit:** ____________________________

***Indicates identifying information which will be used solely to provide extra credit**

Age:
____

Race/Ethnicity:
____

Gender (*Please circle one*): Male Female

LSU Classification (*Please circle one*):

Freshman Sophomore Junior Senior Graduate Student

Have you ever been diagnosed with a neurological disorder (e.g., epilepsy, traumatic brain injury, meningitis, encephalitis, extreme fever, stroke, hematoma)?

Yes No

If Yes, Please explain:
____

Have you ever been diagnosed with a psychological disorder (e.g. depression, anxiety, ADHD, learning disability, OCD, schizophrenia, bipolar disorder, alcohol or substance abuse or dependence)?

Yes No

If Yes, Please explain:
____

Examiner: ____________________________ **Date:** ____________________________

APPENDIX D: PARTICIPANT EFFORT RATING SCALE

Subject Number: ________________________

Examiner: _________________ Date:______________________

How much effort did you put into performing as you think someone who is mentally retarded would? (*Please circle one number*)

1	2	3	4	5
Not much		Average		Very much

How successful do you think you were at performing as someone who is mentally retarded would? (*Please circle one number*)

1	2	3	4	5
Not at all		Average		Very Successful

VITA

Mandi Wilkes Musso was reared in Simmesport, Louisiana by mother Wanda Wilkes, and father Marcus Wilkes. Mandi performed well in school and participated in multiple extracurricular activities. At the age of 13, Mandi decided that she wanted to be a Clinical Psychologist, a career path that she pursued when she began taking classes at the University of Louisiana at Lafayette in 2002.

While at the University of Louisiana at Lafayette, Mandi was an active member of Psi Chi and was elected President in 2005. She also worked as a research assistant at the New Iberia Research Center in New Iberia, Louisiana, completing an Honor's Thesis on the self-injurious behavior in Rhesus Macaques. She graduated Summa Cum Laude with an Honors Bachelors of Science degree in psychology and a minor in biology in 2006. She also received the Hait Lewis Award for Academic Excellence. That same year, she was married to M. Dustin Musso.

After graduating in 2006, Mandi worked full time as a research associate at the New Iberia Research Center while continuing to take classes in biology and chemistry. In 2007, she met with Dr. William Drew Gouvier to discuss his program of research at Louisiana State University and decided that she wished to enroll in that program. Mandi is currently a third-year doctoral student and a student member of the National Academy of Neuropsychology. Her research interests include neuropsychology, police psychology, and malingering assessment.