Improving the Ability of the MMPI-2-RF to Discriminate between Psychogenic Non-epileptic Seizures and Epileptic Seizures

by

Rebecca E. Wershba

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2012 by the
Graduate Supervisory Committee:

Richard I. Lanyon, Chair
Manuel Barrera
Paul Karoly
Roger E. Millsap

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

The use of bias indicators in psychological measurement has been contentious, with some researchers questioning whether they actually suppress or moderate the ability of substantive psychological indictors to discriminate (McGrath, Mitchell, Kim, & Hough, 2010). Bias indicators on the MMPI-2-RF (F-r, Fs, FBS-r, K-r, and L-r) were tested for suppression or moderation of the ability of the RC1 and NUC scales to discriminate between Epileptic Seizures (ES) and Non-epileptic Seizures (NES, a conversion disorder that is often misdiagnosed as ES). RC1 and NUC had previously been found to be the best scales on the MMPI-2-RF to differentiate between ES and NES, with optimal cut scores occurring at a cut score of 65 for RC1 (classification rate of 68%) and 85 for NUC (classification rate of 64%; Locke et al., 2010).

The MMPI-2-RF was completed by 429 inpatients on the Epilepsy Monitoring Unit (EMU) at the Scottsdale Mayo Clinic Hospital, all of whom had confirmed diagnoses of ES or NES. Moderated logistic regression was used to test for moderation and logistic regression was used to test for suppression. Classification rates of RC1 and NUC were calculated at different bias level indicators to evaluate clinical utility for diagnosticians.

No moderation was found. Suppression was found for F-r, Fs, K-r, and L-r with RC1, and for all variables with NUC. For F-r and Fs, the optimal RC1 and NUC cut scores increased at higher levels of bias, but tended to decrease at higher levels of K-r, L-r, and FBS-r. K-r provided the greatest

suppression for RC1, as well as the greatest increases in classification rates at optimal cut scores, given different levels of bias.

It was concluded that, consistent with expectations, taking account of bias indicator suppression on the MMPI-2-RF can improve discrimination of ES and NES. At higher levels of negative impression management, higher cut scores on substantive scales are needed to attain optimal discrimination, whereas at higher levels of positive impression management and FBS-r, lower cut scores are needed. Using these new cut scores resulted in modest improvements in accuracy in discrimination. These findings are consistent with prior research in showing the efficacy of bias indicators, and extend the findings to a psycho-medical context.

ACKNOWLEDGMENTS

First, I would like to give my heartfelt thanks to my adviser, Dr. Richard Lanyon, who helped a health psychology/neuropsychology student find her direction. Your knowledge and wisdom were invaluable, and I always left your office more inspired, motivated, and frankly, entertained than when I entered. Many thanks also to Dr. Dona Locke, my neuropsychology supervisor at Mayo Clinic. Your clinical work and research have served as a major inspiration, and my future career direction is largely due to your influence. You generously provided me with this project and the data to undertake it, and watched it turn into a dissertation. Without you, this study would not have happened. I cannot thank you enough. Much gratitude also goes to my committee, Drs. Manuel Barrera, Paul Karoly, and Roger Millsap (statistician extraordinaire). Your suggestions were always on point and my study is better because of your help.

I must also acknowledge my fellow graduate students who got me through this process in one piece. Your support, commiseration, and the occasional beer made these graduate years fuller and richer, and I appreciate your friendship. To those of you who talked me though statistics, editing, formatting, and all the assorted joys that come with a project of this magnitude: You are lifesavers and part of this paper rightly belongs to you. You may have Table 13.

Finally, thank you to Josh, who provided me with love and support through the whole of this project. You are the only business person in the world who could, as a result of hours and hours of patient listening, explain

RC1 to the casual observer and defend a psychology dissertation. I am unbelievably lucky to have you in my life.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Introduction

Validity scales, used to measure response bias, are commonly used in personality and clinical assessment to gauge how much credibility should be placed in substantive scales. These scales indicate whether a test-taker is approaching the test in a manner which is likely to over-represent somatic or psychiatric dysfunction, under-represent these characteristics, or result in randomness which would render the entire test noncredible; however, this is generally the only utility afforded these very interesting and informative scales. But rather than simply providing information as to whether the clinical variables can be usefully interpreted, validity scales may also hold information as to how the substantive variables should be interpreted. The same score on a depression scale, for example, may hold different diagnostic utility for a person who indicates through his or her response style that he or she is reporting a non-credible level of psychopathology versus a person who seems to be accurately reporting his or her level of functioning. Whereas there is literature suggesting that response bias affects the discriminant utility of clinical scale scores in cases such as psychiatric malingering and work force testing, there is a dearth of such literature in the medical field. At this point, the author is unaware of any studies examining the ways in which validity scales affect medical diagnostic utility of clinical scales.

The present study focused on this gap in the literature by exploring methods of using validity scores to improve the diagnostic utility of clinical scores on personality-related tests in a medical setting. Specifically, the present study tested the ability of bias indicators to either suppress or

moderate the predictive utility of clinical scale scores in medical and psychological diagnoses. In the case of suppression or moderation, predictive success of clinical scale scores at different levels of response bias was calculated to improve diagnostic utility. This method has not been utilized to date in the psycho-medical field.

The present study builds on Locke et al.'s 2010 study of discrimination of epileptic seizures (ES) and non-epileptic seizures (NES), a conversion disorder that mimics seizure activity without an underlying neurological disorder. Differentiating between ES and NES is a useful way to begin to address the problem of improved diagnosis through psychological assessment, as there are consistent differences in both validity and clinical scale scores between these populations, as well as reliable methods of medically diagnosing these cases. If the methods used in the present study succeed in improving diagnostic abilities of personality tests, it will provide a valuable technique that may have implications for diagnosis beyond ES and NES. It is also hoped that, in the process of developing new criteria for the discrimination of ES and NES by utilizing validity scales, this study will increase understanding of psychological processes involving NES. Because the Minnesota Multiphasic Personality Inventory-Restructured Form; (MMPI-2-RF; Ben-Porath & Tellegen, 2008) is in the process of becoming one of the most widely used personality tests, and as it has a range of well-studied validity scales, it was the assessment tool used in this study.

The present review will first define and describe NES and the common characteristics of those that suffer from this unique disorder. It will then

discuss methods of differentiating NES from ES, first through physiological differences in NES and ES, then by psychological assessment strategies. Response biases and the ways in which they affect psychological assessments will be explored, as well as methods of correcting for response biases. Finally, the assessment indicators of greatest potential interest in this population will be described.

Non-epileptic Seizures

NES are a specific form of conversion disorder, defined as the presence of "pseudoneurological" motor or sensory deficits, suggestive of a medical or neurological condition, and in the absence of any known underlying medical condition (American Psychiatric Association Task Force on DSM-IV, 2000). NES is the third of the four subtypes of conversion disorder listed, which are (a) With Motor Symptom or Deficit (such as paralysis, localized weakness, or inability to retain urine), (b) With Sensory Symptom or Deficit (such as anesthesia, blindness, deafness, or hallucinations), (c) with Seizures or Convulsions (NES), and (d) With Mixed Presentation. In all subtypes of conversion disorder, these deficits generally do not follow known anatomical or physiological pathways, the person rarely shows expected objective signs of physiological dysfunction, and presentation is reliant upon the patient's conceptualization of what such deficits may look like.

Of all conversion disorders, NES is unique in that, unlike other conversion symptoms such as pain or weakness, it is it is possible to objectively measure whether the symptoms are being caused by expected neurological activity in the moment. Epileptiform activity can be measured in

the brain during observed seizure activity through the use of 24-hour video monitoring with accompanying electroencephalogram, or video-EEG (VEEG), the "gold standard" of epilepsy diagnosis (Benbadis, 2005a; M. Reuber & Elger, 2003). With rare exception, the presence of such activity during seizure-like motor activity indicates ES, whereas the absence indicates a psychogenic cause. Although it is possible to have both ES and NES, it is generally rare, occurring in 5-10% of patients being referred to epilepsy monitoring units (EMUs) for differentiation of ES and NES (Lesser, Lueders, & Dinner, 1983; Locke et al., 2010; Martin et al., 2003), though in some studies, up to 20%-50% of people had mixed ES/NES (Gates, 2002; Mari et al., 2006; Sigurdardottir & Olafsson, 1998)

NES have gone by many names in their time, including factitious seizures (Schachter, Brown, & James Rowan, 1996). NES, and indeed all conversion disorders, should be differentiated from malingering or factitious disorders: the patient is not consciously faking symptoms for secondary gain, either external/financial, or emotional. But it can be difficult to distinguish between malingering/factitious disorder and conversion, especially when conversion symptoms are unbelievable and implausible (Babin & Gross, 2002). Only intent differentiates a conversion disorder from malingering or factitious disorders, and this can be hard to accurately gauge.

Other names for NES include such terms as psychogenic seizures, psychogenic non-epileptic seizures (PNES), pseudoseizures, pseudoepileptic seizures, and hysterical seizures (Babin & Gross, 2002). Many of these have, for the most part, been rejected as inaccurate or derogatory (Gates, 2002);

however, it has been argued that the term "non-epileptic seizure" is, itself, incomplete. There are seizures that are non-epileptic but physiologic in nature, including those caused by syncope, transient ischemic attacks, acute neurological insults, nontoxic acute hallucinosis, and endocrine disturbances (Gates, 2002). All these can be accurately described as NES. Nonetheless, NES is the most common and preferred nomenclature for NES that are psychogenic in nature.

Conversion disorder is not uncommon, especially in a medical context. Whereas the prevalence of conversion disorders in the general population is estimated to be 11 - 500 per 100,000 people, it is estimated to be 1-14% of the general medical/surgical inpatient population (American Psychiatric Association Task Force on DSM-IV, 2000). Additionally, among almost 4500 neurological inpatients with "typical neurological symptoms", a full 9% were found to have primary causes that were psychogenic in nature rather than neurological (Lempert, Dieterich, Huppert, & Brandt, 1990). For the NES subtype, the prevalence in the general population is estimated at 2-33 per 100,000 people, and between 15-30% of people admitted to EMUs for ES/NES differentiation (Benbadis & Allen Hauser, 2000; Benbadis, 2005b).

Historically, "conversion" implied a literal conversion of anxiety symptoms into a physical symptom or "hysteric phenomenon", such as vomiting in place of moral disgust or anguish (Freud & Breuer, 1895). It is beyond the scope of this review to report on the possible and theoretical psychological underpinnings of conversion disorders, including NES. Nonetheless, although Freud's conversion theories are not applied as literally

today, by current DSM-IV-TR definition a conversion disorder must be preceded by some sort of psychological stressor or conflict. The person may not be aware of the role that such a stressor or conflict has created in his or her life (Babin & Gross, 2002). It has been noted that people with conversion disorders are far more likely to have been sexually or physically abused or neglected as children than comparison groups of the general population (Betts & Boden, 1992; Sar, Akyuz, Kundakci, Kiziltan, & Dogan, 2004) and patients with affective disorders (Roelofs, Keijsers, Hoogduin, Naring, & Moene, 2002). Patients with NES are no exception, and physical and sexual abuse is relatively common in this population (Betts & Boden, 1992). One study found that 59% of such persons had been sexually assaulted or raped as a child or adolescent, and 48% had been physically abused (Bowman & Markand, 1996). This same study found that overall, 88% of subjects had experienced significant trauma over the course of their lives, with 77% suffering sexual abuse or rape and 70% experiencing physical abuse. Trauma as a risk factor as been found in multiple studies (Arnold & Privitera, 1996; Rosenberg, Rosenberg, Williamson, Wolford, & George, 2000) and is associated more strongly with NES than ES (Alper, Devinsky, Perrine, Vazquez, & Luciano, 1993).

NES is a disorder primarily found in females, who account for roughly 75% of diagnoses (Gates, 2002). Younger females tend to be more at risk than older females, with one study finding that the incidence of NES begins to decreases after age 24 (Sigurdardottir & Olafsson, 1998). Additionally, people with NES often have concurrent psychological issues and diagnoses. Common

concurrent diagnoses include affective disorders, PTSD, anxiety disorders, other somatoform disorders, personality disorders such as borderline and mixed personality disorder, and dissociative disorders (Bowman & Markand, 1996; Drake, Pakalnis, & Phillips, 1992; Ettinger, Devinsky, Weisbrot, Ramakrishna, & Goyal, 1999; Sar et al., 2004 [Sar's study included people with conversion disorders, 77% of whom were NES]).

Personality testing has also shown general differences between the NES population and the general population, as well as medical controls. A 2004 study found a higher degree of psychopathology in the NES population compared to patients with epilepsy and general population control subjects (Reuber, Pukrop, Bauer, Derfuss, & Elger, 2004). In this study, cluster analysis on the Dimensional Assessment of Personality Pathology - Basic Questionnaire (DAPP-BQ) found three distinct personality types in patients with NES: (a) a personality type resembling borderline disorder (higher in all four higher-order dimensions [emotional dysregulation, dissocial behavior, inhibitedness, compulsivity]; 51%), (b) an overly controlled personality (increased compulsivity with lower lower-order dimensions such as anxiety, self-harm, narcissism, and conduct problems; 44%), and (c) a personality resembling avoidant personality disorder (increased emotional dysregulation, inhibitidness, and compulsivity; 5%). Cluster analysis using Revised NEO Personality Inventory (NEO-PI-R) scores (Cragar, Berry, Schmitt, & Fakhoury, 2005) yielded a different three clusters; (a) very high neuroticism, low extraversion, low openness, high agreeableness, low conscientiousness, (b) average in all domains, and (c) very high neuroticism, average

extraversion, low openness, low agreeableness, and average conscientiousness. When members of these clusters were examined for accompanying pathology on the MMPI-2, the following personality descriptions for these clusters were suggested: (a) "depressed neurotics" (high on the depression scale), (b) "somatic defenders" (significant somatic "conversion V" profile), and (c) "activated neurotics" (anxiety and not depression). Cragar stated that differences in the results of this and Reuber's study may be due to the use of different measures.

Evidence also suggests that psychopathology differs between people with NES only and mixed NES/ES. One study found that dissociative disorders and PTSD were more common in pure NES, whereas affective and personality disorders were common in both groups (D'Alessio, Giagante, Oddo, & Silva, 2006). Another study found that somatoform disorders aside from NES are more common in the NES-only population, whereas personality disorders are more common among the mixed group (Kuyk, Swinkels, & Spinhoven, 2003).

Although this review does not attempt to explain the reasons between these differences in personality and psychopathology among people with NES and ES, it does note that such differences are acknowledged across studies. Importantly, the presence of psychological differences signals the utility of psychological assessment techniques to help differentiate among people with different diagnoses.

Differential Diagnosis of Non-epileptic Seizures and Epileptic Seizures

Diagnosis through physiology. The overlap of symptoms can confuse the diagnosis of NES and ES. In one EMU, there was a mean delay of 7.2 years from the time of initial NES symptoms and subsequent diagnosis (Reuber, Fernández, Helmstaedter, Qurishi, & Elger, 2002). The misdiagnosis of NES as ES can have devastating effects on a person's life and health. With a diagnosis of ES often comes anti-convulsant medication, with 83% of misdiagnosed patients seen in one EMU receiving anti-convulsant medications during the course of misdiagnosis (Smith, Defalla, & Chadwick, 1999). Anti-convulsants can have a host of side effects, including cognitive deficits (Goldberg & Burdick, 2001), neurotoxic effects, gingival hypertrophy, and gynaecomastia (Smith et al., 1999). Additionally, a misdiagnosis of ES may lead to negative psychosocial outcomes, such as loss of driving abilities (in one study, 12/14 people had their driving interrupted, with five temporarily losing their licenses [Smith et al., 1999]). With loss of driving, or with the stigma of ES, can come interruptions in work or demotion. As such, it is vital that NES be properly diagnosed as early as possible and before these negative effects occur.

Routine EEG (r-EEG) is one method that has been used to help with the classification of epilepsy. It is generally performed on an outpatient basis and involves the gathering of interictal brain activity (Cragar, Berry, Fakhoury, Cibula, & Schmitt, 2002). Epileptiform discharges, even in the absence of seizure activity, is an indication of ES. Strengths of the r-EEG include the relative ease and low expense; however, if a person does not have

a seizure event while being monitored, EEG cannot be used as a discriminating tool (Cascino, 2002). The (possibly rare) co-occurrence of NES and ES may cause a person with mixed NES/ES to be improperly diagnosed by the presence or absence of epileptiform activity, missing the opportunity to medicate true ES or behaviorally treat NES. Additionally, overinterpretation of normal EEG phenomena or non-specific features can lead to a misdiagnosis of epilepsy (Cascino, 2002; Eirís-Puñal et al., 2001; Smith et al., 1999).

VEEG is considered the "gold standard" of epilepsy and NES diagnosis. Lack of epileptiform brain activity during a seizure indicates that there is at least an NES component of a person's syndrome, whereas unusual electrical activity indicates that a person is having an epileptic seizure. VEEG may take place over a number of days, and as such, is more likely to capture multiple events; however, even VEEG can be inconclusive if a person has no seizure activity during the stay. Certain types of seizures, such as frontal lobe seizures, may not show a seizure-like pattern on the EEG (Cascino, 2002). In addition, VEEG may not be available in all locations, and can be a costly venture in both time and money, costing up to $15,000 (Wagner, Wymer, Topping, & Pritchard, 2005). As such, though VEEG is an immensely important tool in the diagnosis of seizures, additional information is usually needed in making a differential diagnosis.

Patients with NES often show seizure activity in a manner consistent with their beliefs about seizures look like, but that are inconsistent with the way seizures physically play out. Observation can therefore be a powerful tool in diagnosis. A summary of differences in seizure activity (Reuber & Elger,

2003) notes common elements in NES that are rare or very rare in ES, including gradual onset, side-to-side head shaking, undulating motor activity, asynchronous limb movements, closed eyelids, eyelids that resist opening, and lack of cyanosis. An additional symptom that has been considered useful in NES and ES differentiation is rhythmic pelvic thrusting, which had been suggested as evidence of a history of childhood sexual abuse in NES. But although this does occasionally occur in NES, it is also an occasional occurrence in frontal lobe epilepsy, though rarely in temporal lobe epilepsy (Geyer, Payne, & Drury, 2000). In people with both NES and ES, clinical features of their seizures generally differ depending on what type of seizure they are having; ES resemble common ES, whereas NES resemble the NES of others (Devinsky, 1996).

Diagnosis through psychological assessment. Another tool that can be used in differential diagnosis is the personality inventory. In addition to giving physicians valuable information into personality and possible psychological dysfunction, knowing how different groups answer such questions can aid in making a diagnosis. One of the most commonly used tests in a hospital setting is the Personality Assessment Inventory (PAI; Morey, 1991). Although this is less studied than other personality inventories, such as the MMPI-2, it has been shown to have great promise in differential diagnosis (Locke, 2011). The PAI is a 344-item test that comprises 22 scales: 4 validity scales, 11 clinical scales, 5 treatment consideration scales, and 2 interpersonal scales. Scales were developed theoretically based on DSM-IV-TR criteria for Axis I and some Axis II

disorders (American Psychiatric Association, 2000), as well as variables that assess openness to treatment, alcohol and drug abuse, and interpersonal scales. *T* scores were developed with a community sample of 1000 individuals.

Wagner et al. (2005) was the first to study the use of the PAI in the EMU. Using VEEG as a final arbiter for diagnosis, differences were found in the SOM-C (Conversion subscale of the PAI Somatization [SOM] scale) between people with NES such that people with NES scored significantly higher (*T* score of 71.8 vs. 68.3). As research had found that SOM-C is greatly increased in people with conversion disorders and the Health Concerns scale is generally increased in people with serious health problems, he additionally created an "NES indicator" of SOM-C minus Health Complaints, with the hypothesis that a positive result would be an indicator of NES, and a negative result would be an indicator of ES. A positive NES indicator yielded a 84% sensitivity and a 73% specificity; however, when replicated with a larger sample (Thompson, Hantke, Phatak, & Chaytor, 2010), the NES indicator was not found to have greater accuracy than the full SOM scale, either of SOM's subscales SOM-C and SOM-S (Somatization subscale), or DEP-P (Physiological Subscale of the PAI Depression scale). Additionally, in Thompson's sample, the mean score on the Health Concerns scale was the same with both groups. A further study by Locke et al. (2011) which compared those scales of the PAI, MMPI-2 and MMPI-2-RF that had shown the most potential for their ability to discriminate ES and NES in an EMU sample found that SOM-C at ≥70 was a better discriminator of NES than

Wagner's NES indicator subscale, the SOM scale, or the MMPI-2RF scale RC1 (Somatic Complaints).

Perhaps the most used and best researched personality test is the MMPI (Hathaway & McKinley, 1940; Locke et al., 2011) and its more recent versions, the MMPI-2 (Butcher et al., 2001) and the MMPI-2-RF (Ben-Porath & Tellegen, 2008). The MMPI was originally developed as a diagnostic classification tool, but it became clear that specific diagnostic utility as planned, i.e., using single high scores on individual scales such as hypochondriasis, depression, and schizophrenia for diagnosis, was poor (for names of relevant MMPI-2 and MMPI-2-RF scales and their abbreviations, please see Tables 1 and 2); however, it provided useful information for psychiatric characteristics through code types and patterns (such as a high Hs/Hy or 1-3 code as a predictor of a psychosomatic disorder; McKinley & Hathaway, 1944), which gained more interest than individual scales. The MMPI was revised in 1989 as the MMPI-2, with new norms, removal of offensive, outdated, and irrelevant questions, and creation of new scales, including the validity scales VRIN, TRIN, and Fb. In 1995, Fp, S, and the PSY-5 scales were added. Review articles (Cragar, Schmitt, Berry, Cibula, Dearth, & Fakhoury, 2003; Locke et al., 2010) indicate that people with NES tend to have elevations on the MMPI-2 scales 1 and 3, and to a lesser extent, 2 and 8. These same reviews found that elevations on scales 1 and 3 are good differentiators of ES and NES, with lesser differentiation found in scales 2 and 4, and nonsignificant differences in scale 8. Since then, decision rules have been created to best distinguish NES from ES. Three such sets of

decision rules are given in Table 3. Wilkus' decision rules for discriminating NES from ES using the MMPI (Wilkus, Dodrill, & Thompson, 1984) were generally based around the "conversion V" phenomena, whereby elevated Hs and Hy scores, which are higher than a lesser-elevated D, are related to conversion disorders (Gough, 1946). These decision rules, when applied to ES and NES groups, result in a sensitivity of 61-80% and specificity of 40-88% (review article; Cragar, Schmitt, Berry, Cibula, Dearth, & Fakhoury, 2003). These rules were later modified by Cragar for the MMPI-2 (Cragar, Schmitt, Berry, Cibula, Dearth, & Fakhoury, 2003; See Table 3). Sensitivities for the modified rules ranged from 58-74%, whereas specificities ranged from 70-74%. A similar set of decision rules was created by Derry and McLachlan (1996) which also capitalizes on the relative elevations of scales 1 and 3 among people with NES (see Table 3). The original paper, which included mixed ES/NES, reported sensitivities of .92 and specificities of .94 using rules created after collecting data to maximize classification accuracy in that particular group. Using the same decision rules, another study found a sensitivity of .71 and specificity of .67 (Warner, Wilkus, Vossler, Wyler, & Abson-Kraemer, 1996).

A shift in the MMPI-2 occurred in 2003 with the addition of the Restructured Clinical (RC) scales as potential replacements of the Clinical Scales, designed to decrease the common variance and intercorrelations among the Clinical Scales. Variance due to demoralization (high negative activation and low positive activation) was removed from all scales to reduce correlations between the scales, and was made into its own scale,

Demoralization. Scales 5 and 0 (Masculinity-Femininity and Social Introversion, respectively) were removed, leaving Demoralization and eight scales which generally correspond to the concepts underlying the original clinical scales. For example, RC1, Somatic Complaints, measures the same general concept as Clinical Scale 1, Hypochondriasis. A major exception is RC3, which measures Cynicism rather than Hysteria, and is in fact a reversal of the naïveté component of Clinical Scale 3. As 1 and 3 were previously highly related, the restructuring removed the intercorrelations such that the somatic elements of Clinical Scale 3 are now associated with RC1, which independently reflects somatic complaints (in contrast to the MMPI-2's "conversion V" of a high 1 and 3; Sellbom, Ben-Porath, McNulty, Arbisi, & Graham, 2006). A high score on RC1 and low score on RC3 is meant to recreate the somatic complaining and naïveté that was the hallmark of Clinical Scale 3.

In 2008, the MMPI-2-RF was introduced. It substantially reduced the number of items from 567 to 338 to improve efficiency and enhance construct validity by removing items with less utility, and utilized many of the same methods that were used to develop the RC scales, such as factor analysis (Ben-Porath & Tellegen, 2008). It is composed of 50 scales (for a list of MMPI-2-RF scales used in this review; see Table 1) and is a heterogeneous collection of scales from previous MMPI versions as well as scales created by various authors and deemed to have optimal utility.

A study on the ability of the MMPI-2-RF to distinguish ES from NES (Locke et al., 2010) found mean differences in the validity scales Fs and FBS-

r, clinical scales RC1 and RC3, and somatic scales MLS, GIC, HPC, and NUC (Persons with NES scored higher on all these scales except for RC3; for descriptions of scales, please see Table 1). FBS-r, RC1, and MLS had the largest effect sizes of their respective classifications. At a cut score of T=65, RC1 had a sensitivity of 76% and a specificity of 60%, with an overall hit rate of 68% (the base rate of NES in the population used in this study, namely, people with VEEG-diagnosed NES or ES, was 50.9%). For NUC, at a clinical cut score of 65, sensitivity was 91 but specificity was only 27; the hit rate was maximized at a cut score of T=85 (sensitivity 53 and specificity of 81; hit rate 67%). Although there was no optimal clinical cut score determined for FBS-r, when FBS-r ≥70, sensitivity was 56 and specificity was 73, with an overall hit rate of 64. Locke & Thomas (2011) also developed two scales for the MMPI-2-RF that measured physical complaints (Psychogenic Nonepileptic Seizures Physical Complaints; PNES-pc) and attitudes (Psychogenic Nonepileptic Seizures Attitudes; PNES-a). These scales were developed through identifying individual questions that differentiate between NES and ES and heuristically sorting them into scales, which were then evaluated and refined through confirmatory factor analysis. When PNES-pc and PNES-a were both ≥3, sensitivity and specificity were both 73%.

A recent study by Locke et al. (2011) sought to compare the best of the decision rules for the PAI (SOM, SOM-C, and NES indicator), MMPI-2 (Derry et al. and Wilkus et al. decision rules), and MMPI-2-RF (RC1, PNES indicator). In this sample, patients in an EMU were randomly assigned to take either the PAI or the MMPI-2, which was additionally scored on the

MMPI-2-RF scales. Of the 78 people who took the PAI, 46 were confirmed by VEEG to have ES, whereas 32 had NES; of the 65 that took the MMPI-2, 33 had ES and 32 had PNES. Between the MMPI-2 and MMPI-2-RF, RC1 ≥65 had the best discriminating ability with a sensitivity of 97%, specificity of 50%, and hit rate of 73% (base rate in the population used in this study was 38%, which included all people with any confirmed diagnosis, including mixed NES/ES and other physiological disorders); however, the PAI indicators SOM and SOM-C both outperformed the MMPI-2-RF, with SOM having a sensitivity of 83, specificity of 77, and hit rate of 79, whereas SOM-C had a sensitivity of 72, specificity of 84, and hit rate of 79.

As documented above, empirical evidence shows that people with NES and ES are different in key ways in both psychopathology and personality. Personality tests, including the PAI, MMPI-2, and MMPI-2-RF, can be used to predict inclusion in these diagnostic categories. Although VEEG is unmatched in its near-certain diagnostic ability, the aforementioned tests are inexpensive and do not require costly and rare machinery or a great deal of time; as such, they are useful early tools when NES is suspected. Importantly, these tests can be used in combination with other relatively inexpensive tests, such as r-EEG, to provide incremental classification ability (Storzbach, Binder, Salinsky, Campbell, & Mueller, 2000). For people without the resources for VEEG, these tests can together provide a picture of what is happening with the patient with minimal cost and without travel to a specialized hospital. Improving prediction can be an important step toward earlier and more accurate diagnosis, potentially preventing patients from

taking unnecessary medications, facing negative sociological consequences, and losing valuable treatment time.

Response Bias

Response bias, as assessed by what have traditionally been termed validity scales, has long been a source of concern for people who create and use psychological assessment procedures. Response bias has been described as a "consistent tendency to respond inaccurately to a substantive indicator, resulting in a systematic error in prediction" (McGrath, Mitchell, Kim, & Hough, 2010), where a "substantive indicator" is the psychological instrument's scale that is designed to predict a criterion (e.g., the substantive scale RC1 may be used to predict a criterion of diagnosis of a somatizing disorder). Response bias may be done in such a way to create an unusually good impression of psychological functioning; this is known as positive impression management (PIM). Conversely, it may reflect functioning that is far worse than would be assessed objectively; this is known as negative impression management (NIM). Both of these response biases are discussed at length in the present review. An additional type of response bias is random responding, which may be a result of lack of attention or understanding, or other factors. It can be measured by asking paired questions that would be expected to be answered either the same way or in opposite ways. An unusual number of disagreements on questions where agreements are expected (e.g. the VRIN scale on the MMPI; "I have had very peculiar and strange experiences" and "I have strange and peculiar thoughts") or not expected (TRIN; "I am a very sociable person" and "I find it hard to make talk when I

meet new people") indicate that the test-taker is not responding to the test in a meaningful way. Randomly responding to substantive questions invalidates substantive indicators and generally, beyond a set cut point of random responses, results in the entire protocol being considered invalid and discarded. As such, random responding will not be considered in this study, and such invalid protocols will not be utilized.

Response bias may reflect either unconscious or conscious processes, analogous to the psychological concepts of conversion disorder versus factitious disorder/malingering. For example, due to unconscious processes, persons with poor self-understanding may show an unusually high amount of PIM, whereas those with conversion or somatizing disorders may show very high NIM. On the other hand, there are many reasons why a person may deliberately wish to distort his or her testing results, such as or pre-employment assessments given in the workplace or child custody evaluations (Bagby, Marshall, Bury, Bacchiochi, & Miller, 2006). Similarly, those who wish to donate kidneys have been shown to have unusually high PIM scores on the PAI (Hurst, Locke, & Osborne, 2010). Overall, it is logical that people may wish to give a better impression of functioning in contexts involving motivation to get a good job, gain or retain custody of a child, or donate organs to friends or family. Research also abounds with contexts where deliberate NIM is a concern. In forensic assessment, people may be motivated to feign psychological disorders, and in worker's compensation cases or litigation, patients may feign somatic and neurological complaints to make a better case for a lawsuit (Bagby et al., 2006)

As there are no tests of response bias that can discern a respondent's underlying motivation, this (if it exists) must be inferred through the use of context. Psychologists must be careful when making conclusions based on either clinical or validity scales. People with unusual symptoms will be more likely to respond in a way that suggests response bias where it may not exist. For example, on the MMPI-2, patients with seizures are likely to score high on both the clinical scale Sc and the validity scale F due to legitimately bizarre symptoms, such as seeing, hearing, or smelling things that do not actually exist. A careful clinician will recognize this context and not assume that the patient is faking; however, a patient with a mild traumatic brain injury (TBI) obtained in a car crash who never lost consciousness and is suing the insurance company should probably be taken with a grain of salt when he endorses these same bizarre symptoms.

Validity scales as discriminators. Many validity (response bias) scales have been extensively studied to determine whether they do, in fact, discriminate between persons who employ NIM or PIM and those who are answering honestly. There are several research designs used in most studies of response bias, elucidated by Rogers (2008). The first is *simulation research*, in which participants are asked to feign responses in the way a malingerer might do (e.g., pretending they have been in a minor car accident and wishing to represent themselves as more injured than they were for litigation purposes). These feigned responses are then compared to honest responders or a clinical group. This type of testing has high internal validity, since it can be carefully standardized and manipulated, but Rogers states

that external validity is weak, since consequences of poor "malingering" are generally less grave than in a real situation. He additionally cautions against research which compares feigning samples with nonclinical samples, as higher scores for clinical reasons may be seen in genuine patients.

In *known-groups research*, studies compare the responding styles of independently verified different groups (e.g., malingerers and non-malingerers). External validity is strong, but internal validity is weak. Although known-groups research is in many ways optimal due to high external validity, there can be difficulty in independently creating known groups, and categorization is rarely 100% accurate, especially when there is an intent on the part of the person being classified to be categorized into a different group than he/she objectively belongs, as in the case of a malingerer. But it should be noted that NES and ES provide an excellent opportunity for known-groups research, since group inclusion can be objectively measured with VEEG. A lesser-used research strategy is *differential prevalence designs*, in which members of one group are assumed to have a different response style than another. For example, in this method, litigants may be assumed to be a proxy for malingerers, since there is overlap in these populations. Rogers recommends caution when using this research design.

In general, studies examine PIM bias indicators (e.g., L and K on the MMPI) separately from NIM indicators (e.g. the F family, including FBS-r, which contains elements of both PIM and NIM) due to the very different populations of interest. NIM is more of a concern in situations where respondents have a motivation to appear worse than they are, such as cases

of malingering or litigation, whereas PIM is more common when respondents are motivated to appear unrealistically virtuous or problem-free, as in pre-employment assessment or child custody litigation.

A 2003 meta-analysis of the utility of NIM indicators on the MMPI-2 (Rogers, Sewell, Martin, & Vitacco, 2003) indicated that both F and Fp (equivalent to Fp-r, see Table 1; the suffix "-r" denotes the scale has been restructured for the MMPI-2-RF) discriminated well for feigning, and that Fp was especially effective. Other studies have confirmed the ability of F and Fp to distinguish between normal participants instructed to malinger and true psychiatric patients (Bagby, Nicholson, Bacchiochi, Ryder, & Bury, 2002), whereas F and F-back (uncommon responses on the second half of the MMPI-2) distinguish feigned depression from true depression (Bagby, Nicholson, Buis, & Bacchiochi, 2000). In the MMPI-2-RF, FBS-r and Fs were found to correlate with failure on symptom validity tests in disability assessment settings (Gervais, Lees-Haley, & Ben-Porath, 2007; Wygant et al., 2009). More recent meta-analyses of the MMPI-2 found that FBS-r was even more effective than the F family in discriminating overreporting in forensic contexts (Nelson, Sweet, & Demakis, 2006; Nelson, Hoelzle, Sweet, Arbisi, & Demakis, 2010).

PIM indices have also found to successfully classify people who underreport from those who respond without bias. Studies have found that L-r and K-r discriminate between those asked to underreport symptoms and those filling out the MMPI-2-RF under standard instructions (Sellbom & Bagby, 2008). An meta-analysis of 16 studies of PIM on the MMPI-2 found

that underreporting responders show differences in validity scale scores by around 1 standard deviation on effect sizes (Baer, Wetter, Nichols, Greene, & Berry, 1995). This analysis found that the participant and methodological variables in the study (e.g., normals vs. patients and uncoached feigning vs. coached feigning) affected how much PIM impacted scores on clinical scales. In these studies, L-r had effect sizes between -.60 (for coached underreporting faking patients vs. normal controls) and 3.07 (normals feigning and normals standard), whereas K-r had effect sizes between -.04 and 2.06 (both for coached underreporting normals vs. normal controls).

Biasing clinical scores. If a response bias reduces the predictive ability of a substantive indicator on a criterion, there are two possible ways it can do so. If response bias acts as a *suppressor*, this means that it artificially depresses the substantive indicator. If a validity scale score is acting as a suppressor, it can be used additively with the substantive indicator to enhance the prediction of a criterion. For example, the K correction on the MMPI was designed to combat an expected suppressor effect of "defensiveness" on the other criteria, with the assumption that a highly defensive person will not be honest and forthright about impaired psychological functioning. The K correction adds points to other scales dependent on how high the K score is. Statistically, the slope of the regression of the criterion on the substantive indicator will remain the same for all levels of validity scale scores, but the Y intercept will change (McGrath et al., 2010; See Figure 1). Thus, in cases of suppression, bias can be "corrected for" by adding or removing points on the clinical scales.

Response bias can also act as a *moderator* by changing the predictive ability of the substantive indicators at lower or higher levels. That is, with a low or average validity scale score, the substantive indicator may predict a criterion with a high degree of accuracy, but as the validity score increases (indicating more response bias), the substantive indicator may lose predictive validity. VRIN and TRIN are good examples of this effect. With a low VRIN (random responding) score, predictive validity may be quite good, but may decrease dramatically the more VRIN rises. Likewise, if a person is consistently employing PIM or NIM, thereby overreporting or underreporting symptoms, the substantive score may not predict the criterion with much accuracy (see Figure 2). Statistically, if moderation is in effect, the regression slope of the criterion on the substantive indicator will change at different levels of validity scale scores such that low response bias will result in a higher slope, whereas high response bias may result in an almost horizontal slope, indicating no predictive value. In these cases, bias cannot be corrected for by adding or removing points, but it is possible that diagnostic utility can be improved by recognizing the differential predictive success at different levels of response bias.

According to McGrath et al. (2010), the best method to test for suppression is to demonstrate that a) the substantive indicator is correlated with the bias indicator and b) the substantive indicator correlates more strongly with the criterion when the bias indicator is partialed out (Cohen, Cohen, West & Aiken, 2003). The latter is indicated when the semi-partial correlation is larger than the zero-order correlation. An equivalent

correlation indicates that there is no suppression effect; a greater correlation may indicate that the bias indicator actually adds predictive value (for example, if positive impression positively correlates with success in the workplace due to a testee's motivation to appear high-functioning). Another method is using hierarchical regression analyses, first entering the substantive indicator alone, and then together with the bias indicator. If the regression coefficient of the substantive indicator increases upon addition of the bias indicator, this is indicative of suppression (Millsap, personal correspondence).

There are at least two ways to measure moderation. One way is to use moderated multiple regression, or moderated logistical regression in the case of a dichotomized criterion (Cohen, Cohen, West & Aiken, 2003). In this procedure, the increase in fit of adding a multiplicative or interaction term to a regression determines whether, and how much, moderation exists. A second, more clinically useful way of measuring moderation is to examine the correlation between the substantive score and the criterion at different levels of the response-bias measure (McGrath et al., 2010). If moderation exists, the correlations will be lower at higher levels of bias indicator, whereas equivalent correlations at different bias indicator levels indicate no moderation (if the correlations were higher at higher levels of the bias indicator, this would indicate an additive effect of the bias indicator).

Although it is generally accepted that bias scores reflect test approach style, there has been some debate in the literature as to whether high validity scale scores actually do affect the utility of clinical scales. Indeed,

McGrath et al.'s (2010) highly contentious recent review article suggests that there is not enough evidence to demonstrate that bias indicators meaningfully affect the relationship between a substantive indicator and a criterion. These authors reviewed a variety of studies involving personality assessment, workplace variables, emotional disorders, eligibility for disability, and forensic assessment. They stated that there were not enough data for drawing conclusions regarding the latter three populations, but for the first two, evidence indicated only mild support for the utility of bias indicators; however, the majority of the studies reviewed tested only moderation *or* suppression rather than both. Evidence presented below suggests that moderation may be seen when there is virtually no overall correlation between the substantive scale and the criterion. One may also question whether the selected response bias scores, substantive scores, and criterion scores relate in such a manner that suppression or moderation would even be expected. As an example, one study included in the review used the Marlowe Crowne Social Desirability Scale as a bias indicator, Perceived Stress Scale as a substantive indicator, and tension physiological measures as criteria. Failure to find a moderating or suppression effect in this case is likely due to scale choice, and this failure may not generalize to bias indicators in general.

A major concern regarding the McGrath et al. (2011) review includes the overly wide-ranging conclusions reached based on the articles reviewed, a concern recently echoed by prominent researchers in the bias indicator field (Rohling et al., 2011). The main populations researched in the McGrath et al.

review, those being assessed in the workplace, are very different from those where criteria involve psychopathology, or where there is strong motivation to distort. Motivation to present well on a workplace assessment may reflect latent personality characteristics that would actually be assets in the workplace, whereas there is little to suggest that impression management in psychopathology or forensic studies reflect actual positive attributes. Additionally, most of these workplace assessments studied PIM and not NIM, which is more commonly a concern in psychopathology. There would seem to be little reason to suspect that these results would relate to areas in which NIM is frequent, either due to unconscious processes (e.g. somatizing) or deliberately (e.g. litigation or forensics). Although McGrath et al.'s suggested statistical methodologies regarding bias indicators appear sound (and in fact, will be used to inform the statistical choices made in the present study), the populations studied and the conclusions reached in their review have limited relevance to the present study of NIM and PIM in a medical setting.

In contrast to the conclusions of McGrath et al., a recent study by Lanyon, Goodstein, and Wershba (unpublished) examined the use of "good impression" in making predictions from personality measures and reported varied findings. Out of four studies on workplace prediction, two found equivalent prediction across good-impression score levels; however, two found no overall correlations between the predictor and criterion, but found significant correlations at lower levels of good-impression. Three further studies involving the assessment of psychopathology found overall stable correlations between the predictor/criteria, but at extreme levels of the good-

impression scale, found significantly smaller correlations. These data indicate that studies that merely report correlations may be missing out on valuable data, namely a moderating effect.

Several studies examined the effect of NIM on scales assessing clinical characteristics. One study examining the effect of feigning on clinical scores (Burchett & Ben-Porath, 2010) examined how simulated psychopathology and somatizing on the MMPI-2-RF affected clinical scales. As expected, simulators did show elevations in overreporting scores. Simulators also showed elevations on related clinical scales and smaller correlations between substantive scales and related criteria than those measured under standard instructions. Another study found that students asked to feign schizophrenia showed increases in F, and also showed elevated Clinical Scales 6 and 8 when compared to true schizophrenics (Bagby et al., 1997). A study on the MMPI-2-RF examined differences between known malingerers of traumatic brain injury (TBI) from non-malingerers and found increases in both FBS-r and the MMPI-2-RF somatic scales (Youngjohn, Wershba, Stevenson, Sturgeon, & Thomas, 2011). The results of these studies provide evidence that increases in NIM response bias scores correspond with increases in clinical scales. Unfortunately, in these studies, suppression and moderation effects were not tested, so it is unknown whether (or how) these biases affected the discriminant utility of the clinical scales.

Among these studies of the impact of validity scales on clinical scales, there are no data from non-forensic psycho-medical settings. It is possible that the findings of studies utilizing forensic or workplace populations do not

generalize to a general medical setting, especially in a conversion disorder where negative or positive impression are considered to be unconscious rather than conscious processes. The present study will begin to fill in this gap in the literature.

Correction for bias. In addition to having response bias scales, some tests (e.g. the MMPI-2; the Sixteen Personality Factor Questionnaire, Fourth Edition [16PF4; Cattell, Eber, & Tatsuoka, 1970], and the Employment Inventory [Paajanen, Hansen, & McLellan, 1993]) also include methods of correcting the assumed bias. For example, if a person's answering style indicates that he or she is employing PIM, a bias corrector may artificially increase some clinical scores under the assumption that, had the respondent not been underreporting his or her symptoms, his or her score would have been higher. Debate regarding the impact of validity scales on clinical scales aside, in general, practitioners tends to favor the use of bias correctors if they are offered. A survey of 36 researchers who authored publications regarding the use of validity scales in industrial-organizational psychology indicated that 56% used tests which had bias correctors and chose to utilize them, 22% used tests that had bias correctors but did not utilize them, 14% indicated that their tests did not have bias correctors but that they would use them if they were available, and only 8% did not have bias correctors and would not choose to use them (Goffin & Christiansen, 2003). Perhaps the best known bias correction is the K correction on the MMPI and MMPI-2. But currently it is not highly regarded, as studies indicate that it does not function

successfully as intended as a suppressor (Barthlow, Graham, Ben-Porath, Tellegen, & McNulty, 2002).

Classification Rates, Sensitivity and Specificity, and Positive Predictive Power and Negative Predictive Power

Although the aforementioned methods for detecting suppression or moderation have been documented in terms of statistical analyses, they do not provide immediate practical clinical utility for diagnosticians. One method for translating this information for clinicians is finding cut scores on clinical scales that best differentiate between members of a target group and people who are not members. At any given cut score, accuracy of diagnosis can be expressed by determining *classification rates,* or the percentage of people accurately diagnosed by the test. Additional information is provided by determining *sensitivity* and *specificity*. Sensitivity is the likelihood that a member of the target group will be correctly identified, and specificity is the likelihood that a person that is not a member of the target group will be correctly identified. There is a direct tradeoff between sensitivity and specificity: increasing the ability to recognize target group members by casting a wider net will cause more nontarget members to be incorrectly included in target group membership, and vice versa. *Positive Predictive Power* (PPP) is the proportion of people positively diagnosed by the test in question that actually have the condition, whereas *Negative Predictive Power* (NPP) is the proportion of people negatively diagnosed that are actually absent of the condition.

Bias Indicators of Interest

Positive Impression Management. The MMPI scales L and K (L-r and K-r in the MMPI-2-RF) are both indicators of PIM. These scales measure claims of unrealistically good functioning and a general denial of psychopathology. Specifically, L-r measures "uncommon virtues", or denial of minor shortcomings and faults that most people would admit to (e.g., "At times I feel like swearing"), and is composed of 14 items. The *MMPI-2-RF Manual* recommends that a *T*-score ≥80 is unlikely even for those from very traditional backgrounds, and that any absence of clinical elevations in such persons is not interpretable. K-r is built of 14 items and measures psychological defensiveness, and is referred to as "adjustment validity", or how well adjusted a person presents himself/herself (e.g., a "false" scoring of "I find it hard to make talk when I meet new people.") The *Manual* stresses that people who truly are well adjusted may also score highly on the K-r scale, and that extra-test information should be utilized when determining whether the K-r score may be an indication of PIM. But a *T*-score of ≥70 is stated to be uncommon even in very well-adjusted people, and underreporting should be strongly suspected such that a lack of clinical elevations is not meaningful .

Negative Impression Management. The MMPI-2-RF scales F-r and Fs are both indicators of NIM. Rather than general psychological adjustment, these scales measure an unusually high level of pathology. F-r ("infrequent responses") is composed of 32 items and is associated with a broad range of psychological, cognitive, and somatic complaints (e.g. "Evil spirits possess me

at times"; "I get anxious and upset when I have to make a short trip away from home.") Even if a person has many complaints in one category, it is unlikely that a person would legitimately have complaints among all those categories, so that a high F-r is a reasonable indicator that a person is overreporting symptoms of pathology. A *T*-score of ≥79 is said to be an indication of possible overreporting of psychological dysfunction, whereas a *T*-score of ≥120 is highly questionable at best. At this level, malingering is strongly suspected, and clinical scales are not interpretable. The Fs scale is composed of 16 items and measures somatic complaints that are infrequent even among verified patients with physical ailments (e.g., "There seems to be a lump in my throat much of the time.") Like F-r, these items cover a wide variety of complaints such that a person with true somatic complaints would be unlikely to endorse enough items to greatly increase Fs. At a *T*-score of ≥81, exaggeration is suspected, though it may be possible for those with medical complaints. A *T*-score of ≥100 is very uncommon even for those with genuine medical conditions, and scores of the somatic scales should be interpreted with caution.

The Symptom Validity-revised (FBS-r) scale. FBS-r is an example of both PIM and NIM, as it involves presenting as both virtuous and impaired. It was created specifically to measure malingering in personal injury litigation, and is described by its author as indicative of goal-directed behavior, whereby the patient is motivated:

> (1) to appear honest, (2) to appear psychologically normal except for the influence of the alleged cause of injury, (3) to

> avoid admitting pre-existing psychopathology, (4) where pre-existing complaints are known or suspected to have been disclosed to the examining clinician or likely to be disclosed to judge or jury, to attempt to minimize those complaints, (5) to hide pre-injury behavior which is antisocial or illegal or to minimize this if it appears that the behavior will be discovered independently, (6) to present an extent of injury or disability within perceived limits of plausibility [these limits vary widely], (7) and related ends. (Lees-Haley, English, & Glenn, 1991)

It was developed "rationally on a content basis" by its author using the original MMPI responses and observations of known malingerers. As the scale was quite specific in its goals to identify malingerers, it was originally called the Fake Bad Scale. It has rightly been pointed out that, by calling it this, it implies motivation for all who score highly and is a prejudicial term; it has since been renamed Symptom Validity. Meta-analyses indicate that FBS is the best MMPI-2 validity scale at differentiating between forensic overreporting groups and comparison groups (Nelson et al., 2006). Though it was not intended for use other than malingering, it differentiates well between other populations, such as ES and NES. This may relate to the common characteristics of malingerers and conversion disorders on psychological testing; e.g., the MMPI-2 "conversion V" and the MMPI-2-RF elevations in RC1 and the somatic scales (Youngjohn et al., 2011), although it is again cautioned that the present study is not attempting to explain the

psychological characteristics that define persons with NES. There is a considerable overlap between the FBS and the MMPI-2 clinical scales associated with conversion; i.e., Hy (32.6%) and Hs (30.2%; Nelson et al., 2006). The FBS-r is composed of 30 items. According to the *Manual*, at T>80, overreporting of symptoms should be suspected. T≥100 is very rare even among those with true somatic ailments, and the somatic and cognitive scales should be interpreted with extreme caution.

Present Study

The present study examined the potential role of bias indicators to affect the ability of substantive indicators to discriminate between ES and NES, either through moderation or suppression. It took two different approaches to this question: one purely psychometric, and the other more practical. The former method utilized statistical approaches to find the presence or absence of moderation or suppression. Finding either of these would lend support to studies indicating that bias indicators do impact outcomes by affecting other, substantive indicators, while refuting critical articles such as McGrath et al. (2010). The second, clinical method assessed whether the ability of substantive scores to diagnose ES or NES changed at different levels of bias, and did so by examining overall classification rates, as well as how sensitivity/specificity and PPP/NPP changed at these different optimal classification rates. It was possible to have statistical significance without any real clinically significant changes in predictive ability of substantive scales. It was also possible that overall optimal classification rates would remain the same; i.e. the optimal cut score may have been

similar across levels of bias, but the degree to which these cut scores at different levels of bias overdiagnosed ES while underdiagnosing NES and vice versa may have changed. For example, the same cut score at different levels of bias could have had similar classification rates, but PPP could have increased and NPP decreased at higher bias, which would be important information for a clinician who is more concerned with catching all patients with ES (even if more people are incorrectly diagnosed with ES rather than NES), or vice versa. These results were specific to this ES/NES population, but the conceptual usefulness may generalize to other psychomedical populations

If moderation was found, this study would then determine the range of bias indicator scales within which useful discrimination can be made. In the case of either moderation or suppression, this study would identify the best substantive scales for correctly discriminating patients with NES from ES, as well as optimal cut scores on those scales for maximum discrimination, at different levels of bias indicators. It would also examine sensitivity/specificity and PPP/NPP at different levels of bias.

As previously stated, the present study builds on the previous work of Locke et al. (2010). The choices of substantive scales and bias indicators in the present study were based on a review of research and on an empirical, rather than theoretical, basis. Although there is certainly room to offer possible conceptual reasons for psychometric differences that have previously been found between patients with ES and NES (e.g., RC1 measures somatizing, which is directly related to NES as a disorder), this study took a

decidedly practical approach to an important practical problem by examining the scales that have been found to best differentiate between ES and NES.

This study examined the effect of NIM indicators F-r and Fs, and the underreporting indicators L-r and K-r, and FBS-r (which contains elements of both NIM and PIM) on the clinical scales RC1 and NUC for the purposes of discriminating between patients with NES and ES. F-r, Fs, and FBS-r were chosen to test NIM because in the present sample, they are significantly elevated in patients with NES (as is Fs for patients with ES; Locke et al., 2010), making these scales of interest in this population. L-r and K-r were not elevated in this study, but were chosen to test the how PIM indicators might bias clinical indicators. NUC and RC1 were selected for this study because they showed the greatest clinical utility in discrimination in this sample.

Hypotheses: General Formulation

It was hypothesized that bias indicators would work in either a moderating or suppressing fashion to affect the ability of substantive indicators (i.e. RC1 and NUC) to discriminate between ES and NES. For either moderation or suppression, use of bias indicators will be able to improve diagnosis by adjusting the optimal cut score and diagnostic sensitivity and specificity of clinical scales based on response bias scores used by clinicians. The presence of moderation would indicate that the differential diagnostic utility of substantive scales (i.e. RC1 and NUC) changes as bias increases or decreases. The presence of suppression would indicate that the optimal cut score on substantive scales changes at different levels of bias, but

diagnostic utility at optimal scores would remain similar across different levels of bias.

Negative Impression Management. The indices of negative impression management assess biased responses to questions about physical or cognitive symptomatology, including the type associated with NES. It was predicted that as patients responded in a more biased manner, higher cut scores on substantive scales (i.e. RC1 and NUC) would be needed to attain optimal differentiation between ES and NES.

Additionally, if moderation was seen, it was predicted that as patients responded in a more biased manner, the differentiating diagnostic utility of substantive scales would decrease, such that sensitivity and specificity at different cut scores on substantive scales would change at different levels of F-r and Fs. In this situation, NIM indices would begin to affect clinical scales substantially as these scales approached *T* scores which the MMPI-2-RF *Manual* suggests indicate possible malingering, i.e. F-r =80 and Fs =81. When there was no indication that there is a biased response style at work, validity indicators would not substantially affect clinical scales.

Positive Impression Management. It was predicted that indices of positive impression management would also affect clinical scales. Underreporting indices such as L-r and K-r do not directly measure symptoms, but rather an approach to answering questions in a defensive or honest style of responding. Since this response style is less directly related to NES symptoms, it was predicted to have less impact on differential diagnosis than NIM and FBS-r. The K-r scale, which was found to be significantly

higher in the NES population than the ES population in Locke et al.'s sample (although this level still did not rise to clinical significance), would affect sensitivity and specificity at different cut scores more than L-r, which did not discriminate between ES and NES in this sample.

If moderation was seen, these scales would affect clinical scales at both high and low levels. At high scores, it was predicted that discriminating ability of clinical scales would be attenuated due to a lack of openness on answering questions. A very low level of PIM can be a result of unusual honesty and openness or can be a result of endorsing an unusual amount of dysfunction (and therefore appearing unusually open). As a result, no hypotheses are made as to whether low levels of K-r and L-r would increase or impair predictive utility of clinical scales.

The Symptom Validity-revised (FBS-r) Scale. FBS-r, as a bias indicator, is not historically used for any diagnostic discrimination, including medical diagnostic discrimination. Rather, it is used as an indicator of whether substantive scales, such as RC1, can be trusted to usefully discriminate in cases of personal injury litigation malingering. However, in this sample and others, FBS/FBS-r itself discriminated between NES and ES such that persons with NES scored significantly higher (Barr, Larson, Alpert, & Devinski, 2005; Locke et al., 2010; Nelson, Parsons, Grote, Smith, & Sisung, 2006). This may be due to the item content in FBS, which, as previously stated, overlaps with the somatizing MMPI-2 scales Hy and Hs. Therefore, it was predicted that as FBS-r scores increased, higher cut scores

on substantive scales would be needed to attain optimal differentiation between ES and NES.

In the case of moderation, as FBS-r increases, it was predicted that substantive scales would be increasingly biased and less able to discriminate. It would begin to affect clinical scales substantially as FBS-r approaches *T* scores which the MMPI-2-RF *Manual* suggests indicate possible malingering, i.e. FBS-r =81. As it is itself a predictor of NES or ES, if moderation is seen, at a certain FBS-r score (to be determined), the FBS-r score itself would be the best predictor of inclusion into the NES or ES group.

Specific Hypotheses

In summary, given the potential for response bias to affect a clinical score, the hypotheses were as follows:

Negative Impression Management.

1a) Indicators of NIM (i.e. F-r and Fs) will moderate or suppress substantive scales (optimal predictors RC1 and NUC).

1b) In the case of moderation, at higher levels of NIM, diagnostic utility will be attenuated. In the case of suppression, diagnostic utility at optimal cut scores will be similar across levels of bias.

1c) For either moderation or suppression, the optimal cut score for discrimination between NES and ES will increase at higher levels of NIM.

Positive Impression Management.

2a) Indicators of PIM (i.e. K-r and L-r) will moderate or suppress substantive scales.

2b) In the case of moderation, there will be a moderating effect of the substantive scales on the criteria at both high and low levels. At high levels of PIM, diagnostic utility of clinical scales will be attenuated. No hypotheses are made as to the directionality of how low scores on underreporting indices will affect clinical scales. In the case of suppression, diagnostic utility at optimal cut scores will be similar across levels of bias.

2c) For either moderation or suppression, the optimal cut score for discrimination between NES and ES will decrease as PIM increases. In the case of suppression, diagnostic utility at optimal cut scores will be similar across levels of bias.

2d) PIM will affect clinical scales less than NIM.

The Symptom Validity-revised (FBS-r) Scale.

3a) FBS-r will either moderate or suppress substantive scales.

3b) If moderation is seen, at higher levels of FBS-r, diagnostic utility will be attenuated. In the case of suppression, diagnostic utility at optimal cut scores will be similar across levels of FBS-r.

3c) The optimal cut score for discrimination between NES and ES will increase as FBS-r increases.

3d) If moderation is seen, at a certain FBS-r score, the predictive utility of FBS-r alone will be higher than that of the substantive scales RC1 and NUC.

Method

Procedure

All participants were inpatients in the EMU at the Mayo Clinic Hospital in Phoenix, Arizona between April 2001 and April 2009. A total of 664 patients were admitted (14 of these were second admissions during the same time period and were not included in the final results), and all were given the MMPI-2 as part of a standard neuropsychological evaluation (recoded later into the MMPI-2-RF). Patients were weaned off antiepileptic medications and were monitored with VEEG, where epileptiform discharges were recorded simultaneously with video monitoring. Diagnosis was determined from VEEG discharge summary and was made according to the following criteria:

Epilepsy only: Typical events occurred with epileptiform discharge, or, in the absence of the occurrence of typical events, the description of a typical event was concerning for epilepsy and interictal epileptiform discharges were noted.

NES only: Typical events occurred without epileptiform discharge. There was no interictal epileptiform discharge, and no physiological reason for the seizure-like events.

Both Epilepsy and NES: Typical events occurred multiple times, some with and some without epileptiform discharges, or typical events occurred without epileptiform discharges but interictal epileptiform activity was noted.

Indeterminate: No typical events nor interictal discharges were recorded, resulting in no diagnosis.

Participants

Of the 664 patients, 221 were diagnosed with epilepsy, 219 were diagnosed with NES, 24 were diagnosed with both ES and NES, 166 were indeterminate, and 34 patients were diagnosed with other physiological disorders such as sleep, autonomic nervous system, or vascular disorders. Patients other than pure ES or NES were excluded from inclusion in the study. After excluding the re-admissions among the NES and ES patients, the study was left with 215 ES and 214 NES patients.

Record review gathered the following information: age, sex, handedness, ethnicity, length of disorder, years of education, seizure frequency, current number of anti-epileptic medications being taken, current psychiatric medications, psychiatric history (broadly defined to include a diagnosis of a psychiatric issue, self-report of mood or psychiatric problems, history of individual, family, or marital counseling, evaluation from a psychiatrist, medication for a psychiatric issue, or inpatient psychiatric hospitalization) and substance abuse history (see Table 4 for descriptives).

Basic Statistical Analyses

Moderation. To determine whether response bias indicators act as moderators, two methods were used. The first method utilized moderated

logistical regression. Moderation, in this instance, occurred if the odds of a correct diagnosis increased when an interaction term (response bias score multiplied by clinical scale score) was included in the model. This method is statistically powerful since it allows all independent variables to remain continuous. A second method of measuring moderation examined the correlation between clinical scale scores and diagnosis at different levels, or cut slices, of the response-bias measure. Although less powerful, this method might pick up on moderation that only occurs at extremes, or very high/very low levels of validity scale scores.

Suppression. To determine whether response bias indicators act on substantive indicators as suppressors, binary logistic regression was used with and without the bias indicator as a covariate, whereby X1 =a substantive indicator, X2 =a bias indicator, and Y =VEEG-confirmed diagnosis (e.g., X1 =RC1, X2 =FBS-r, and Y=NES or ES). If the bias indicator is significant, this indicates that it has an additive effect to the substantive indicator for diagnosis. In addition, if the standardized regression coefficient of the substantive indicator is greater when the bias indicator is added, the bias indicator acts as a suppressor, since by being partialed out in the model, the relationship between the substantive indicator and the criterion becomes stronger. Each of the two substantive indicators and the five bias indicators were tested.

Analyses with Clinical Utility: Classification Rates, Sensitivity and Specificity, and Positive Predictive Power and Negative Predictive Power. In addition to testing moderation and suppression, an overall correct

classification rate for different cut scores at different levels of bias can be calculated. Unlike statistical tests of moderation and suppression, this can offer immediate clinical utility for those using the MMPI-2-RF to aid in diagnosis of ES or NES. It can provide information about the likelihood that, at or above a particular cut score, a person will be accurately classified into a target or nontarget group (in this case, the target group is ES and the nontarget group is NES). Classification rates are determined by choosing a cut-score on the substantive indicator and calculating the number of people at or above that score that are correctly identified as members of the target group or non-target group divided by the total sample.

Sensitivity and specificity were also calculated at relevant cut scores of the substantive indicators for different levels of bias indicators. Like classification rates, sensitivity and specificity offer clinical utility for diagnosticians using the MMPI-2-RF. Sensitivity and specificity provide additional information to clinicians about the ability of a particular test score to accurately sort a person, and whether inaccuracies are due to inappropriately sorting people into the target group or the non-target group. Sensitivity is determined by choosing a cut-score on the substantive indicator and calculating the number of people at or above that score that are correctly identified as members of the target group (true positives) divided by the total number of people in the target group (true positives plus false negatives). Specificity is similarly determined by calculating the number of negatives accurately sorted into the nontarget group (true negatives) divided by the

total number of members of the nontarget group (true negatives plus false positives).

Another method of garnering useful information from the data is to look at positive predictive power (PPP) and negative predictive power (NPP). PPP is the proportion of people positively diagnosed by the test in question that actually have the condition, whereas NPP is the proportion of people negatively diagnosed that are actually absent of the condition. PPP and NPP are thus useful for deciding how likely it is that a person above a certain cut score is accurately or inaccurately diagnosed. PPP calculation is true positives / (true positives + false positives), whereas NPP is true negatives / (true negatives + false negatives). PPP and NPV has an additional useful aspect in that it can be calculated at different base rates (BR) than the sample if sensitivity and specificity are known; if doing this, PPP =(BR x sensitivity) / (BR x sensitivity) + [(1-BR) x (1-specificity)], and NPP =1-BR) x specificity / [(1-BR x specificity) + [BR x (1-sensitivity)].

Classification rates, sensitivity, specificity, PPP, and NPP were calculated at a number of substantive cut scores (e.g., RC1 score of $T \geq 50$, ≥ 55, ≥ 60, ≥ 65...) separately for people at different levels of bias indicators (e.g., F-r scores of T =0-49, 55-64...). In addition to providing information about the likelihood of a particular test score to accurately sort into diagnostic groups, the optimal cut scores of substantive indicators for accurate classification of diagnosis can be determined for people expressing different levels of bias indicators. To ensure adequate power, minimum sample size of each slice was determined according to methods espoused by Buderer (1996) using a

confidence interval of 95%, a confidence interval width of ± 10 for sensitivity and specificity scores (e.g., if sensitivity was found to be 80, it is 95% certain that the actual sensitivity falls between 70 and 90), and an expected minimum sensitivity or specificity of 65%. This yielded a minimum sample size per slice of 89.

Results

Basic Statistical Analyses

Correlations between each of the substantive indicators, bias indicators, and the criterion were performed, and are shown in Table 5. Bias indicators were minimally related to the criterion (NES vs. ES), with all these variables being correlated at ≤0.15, with the exception of FBS-r, which was correlated at 0.32. This indicates that, except for FBS-r, bias indicators are not acting as independent predictors of the criterion.

Moderation. The presence of moderation was tested through two methods.

Binary logistic regression. Moderated logistic regression yielded no significant moderating effects of bias indicators for either RC1 or NUC. Power to detect effects, calculated using G*Power, was calculated at 0.97 (Hsieh, Block, & Larsen, 1998; Millsap, personal communication, 2012; Buchner et al., 2009). Results are shown in Table 6.

Correlations between substantive indicators and diagnosis. To test for moderation via correlation of clinical scale scores and diagnosis at different slices of the response-bias measure, bias indicators were divided into three or four different slices. Number of slices, as well as cut scores at slices, was

determined by adherence to the *Manual* guidelines of meaningful cut scores (when possible), and otherwise, attempting to create the maximum number of groups while maintaining adequate sample size in each group. Because *T* scores were not evenly distributed, this led to some bias indicators being split into three groups and others into four groups. This has precedence in the *Manual*; likely due to the same issues, the *Manual*-suggested cut scores of clinical importance are not congruent between bias indicators, nor are the ranges of scores within each clinically important slice within in a single bias indicator the same.

Correlations of substantive indicators and the criterion were performed at different levels of bias. These correlations were then compared to each other using a Fisher r-to-z comparison, using an internet application (Lowry, R., 2001-2012). Power analyses calculated using G*Power indicated that the sample size was insufficient to detect any differences between groups (Buchner et al., 2009); maximum power attained in comparing correlations was only .25. In fact, for a power of 0.8, 3146 participants would be required. Nonetheless, analyses were run as an exercise. Results, shown in Table 7, indicate that none of the correlations were significantly different from each other for either RC1 or NUC. No moderating effect of bias indicators was demonstrated, but it is again warned that these analyses were insufficiently powered.

Suppression analyses for RC1. Suppression was tested through hierarchical binary logistic regression, whereby the substantive variable and each bias variable were individually entered in a sequence to test how much

the bias variable added to the model. Each analysis was composed of RC1 and one only bias indicator. All suppression analyses are shown in Table 8.

RC1 alone. RC1 was first entered alone in binary logistic regression model (B =0.069, s.e =0.009, *p*=.000).

Addition of bias indicators. F-r was added into the original model and was significant (B= -0.031, s.e. =0.009, *p*=.000). The regression coefficient for RC1 increased (B =0.094, s.e.=0.012, *p*=.000).

Fs was added into the original model and was significant (B =-0.024, s.e. 0.008, *p*=.000). The regression coefficient for RC1 increased (B =0.091, s.e. =0.012, *p*=.002).

L-r was added into the original model and was significant (B =0.031, s.e. 0.011, *p*=.000). The regression coefficient for RC1 increased (B=0.076, s.e. =0.009, *p*=.000).

K-r was added into the original model and was significant (B =0.068, s.e. = 0.012, *p*=.000). The regression coefficient for RC1 increased (0.089, s.e. =0.010, *p*=.000).

FBS-r was added into original the model and was not significant (B =0.000, s.e. =0.012, p =.970). The regression coefficient for RC1 stayed the same (B=0.069, s.e., 0.012, *p*=.000).

The finding that the addition of bias indicators F-r, Fs, L-r, and K-r increased RC1's beta coefficient demonstrated that each of these bias indicators suppressed RC1's ability to discriminate between ES and NES ; however, the addition of FBS-r did not increase or decrease RC1's beta coefficient, indicating that it did not suppress RC1.

Suppression analyses for NUC. Suppression was tested in the same way as was done for RC1. Results of suppression analyses are shown in Table 8.

NUC alone. NUC was first entered alone in binary logistic regression model (B=0.065, s.e.= 0.008, *p*=.000).

Addition of bias indicators. F-r was added into the original model and was significant (B =-.023, s.e. =.010, p =.004). The regression coefficient for NUC increased (B =.081, s.e. =.010, p =.000).

Fs was added into the original model and was significant (B =-.017, s.e. .007, p =.023). The regression coefficient for NUC increased (B =.078, s.e. =.011, *p*=.000).

L-r was added into the original model and was significant (B =.031, s.e.=.011, p =.005). The regression coefficient for NUC increased (B=.071, s.e.=0.009 *p*=.000).

K-r was added into the original model and was significant (B =.048, s.e.=.012, p =.000). The regression coefficient for NUC increased (B =0.74, s.e =.009, *p*=.000).

FBS-r was added into the original model and was significant (B =.024, s.e. =.009, p =.009). The regression coefficient for NUC decreased (B=.052, s.e. =0.010, *p*=.000).

The finding that the addition of bias indicators F-r, Fs, L-r, and K-r to NUC increased NUC's beta coefficient indicates that these bias variables suppressed NUC's ability to discriminate between ES and NES. Addition of FBS-r decreased rather than increased NUC's beta coefficient. Since FBS-r

was also independently significant, this indicates that FBS-r can be used additively with NUC to improve prediction of the criterion, but it does not do so through suppressing NUC.

Analyses with Clinical Utility

Classification rates. After assessing for moderation or suppression, analyses for clinical utility were run. To determine classification rates and sensitivity/specificity, bias indicators were each split into either three or four groups of similar size (minimum *n* of 89) depending on the *MMPI-2-RF Manual* suggested cut scores and distributions of *T* scores within each indicator. Based on these criteria, F-r's slices were *T*=0-54, 55-64, 65-78, and ≥79; F-s's slices were *T*=0-59, 60-79, and ≥80; L-r's slices were *T*=0-49, 50-59, and ≥60; K-r's slices were *T*=0-44, 45-54, and ≥55; and FBS-r's slices were *T*=0-54, 55-64, 65-79, and ≥80.

RC1 and NUC were then sliced at scores of *T*≥50, ≥55, ≥60, ≥65... ≥95. For each level of bias (e.g. for F-r, *T*=0-54), all persons above each substantive cut score, (e.g. for RC1, *T*≥50), were predicted to be diagnosed with NES. These predictions were then compared to the number of persons that, at or above that cut score, were actually diagnosed with NES via the gold standard of VEEG. Those that were predicted to have NES and were actually diagnosed with NES were considered to be true positives; those that were predicted to have NES but were diagnosed ES were false positives; those predicted to have ES but were diagnosed with NES were false negatives, and those both predicted to have and diagnosed with ES were considered true negatives. Sensitivity, specificity, and classification rates

were determined using the number of people in each category as described previously. Ten such sets of analyses were conducted, one for each predictor/bias-indicator combination. Table 9 shows one of these analyses, for RC1 and F-r.

As this study is concerned with the manner in which NIM, PIM, and FBS-r may affect substantive indicators, and as trends for these bias indicators were similar for RC1 and NUC, classification trends will be reported by bias indicator and not substantive indicator.

Negative Impression Management: F-r and Fs. For both RC1 and NUC, optimal cut scores as determined by overall classification rate tended to increase as NIM increased. For RC1, at F-r <55, optimal classification rate of 69% occurred at $T \geq 60$ and 65, whereas at F-r ≥79, optimal classification rate of 67.5% occurred at $T = 75$. NUC showed some variability, with a higher cut score for F-r <55 (optimal classification rate of 67.0% occurred at $T = 85$), than for mid-ranges of F-r, which otherwise followed the trend of requiring higher cut scores at higher levels of bias for optimal classification rates (i.e., $T \geq 75$ for F-r =56-64 and $T \geq 85$ for T =65-78). For F-r ≥79, optimal classification rate of 72.2% occurred at $T = 90$.

This trend also occurred for F-s. For RC1, at F-s <60, optimal classification rate of 69.8% occurred at T ≥60, whereas for F-s ≥80, optimal classification rate of 68.1% occurred at $T \geq 70$ and $T \geq 75$. For NUC, at F-s <60, optimal classification rate of 64.7% occurred at $T \geq 80$, whereas for F-s ≥80, optimal classification rate of 73.2% occurred at $T \geq 85$.

Positive Impression Management: L-r and K-r. For L-r and K-r, the opposite tendency was seen; at higher levels of bias, cut scores for optimal classification rate occurred at lower levels of bias. For RC1, at L-r <50, optimal classification rate of 66% occurred at *T*≥75, whereas at L-r ≥60, optimal classification rate of 75.8% occurred at *T*=60. For NUC, at for L-r <50, optimal classification rate of 74.1% occurred at *T*≥85, whereas for L-r ≥60, optimal classification rate of 70.2% occurred at *T*≥70.

Similarly, for RC1, at K-r <45, optimal classification rate of 71.0% occurred at *T*≥75, whereas at K-r ≥55, optimal classification rate of 76.0% occurred at *T*≥60. For NUC, at K-r <45, optimal classification rate of 72.5% occurred at *T*≥90, whereas for K-r ≥55, optimal classification rate of 71.9% occurred at *T*≥70.

FBS-r. Optimal classification rates for different levels of FBS-r were not calculated for RC1 because no suppression or moderation was found. For NUC, the trend for FBS-r was similar to that of L-r and K-r; at FBS-r <55, optimal classification rate of 75.8% occurred at *T*≥70, whereas at FBS ≥80, optimal classification rate of 69.4% occurred at below at or below *T*≥55. At those *T*-scores, that optimal classification rate has a sensitivity of 100% and a specificity of 0%; that is, the best overall classification rate occurs when it is predicted that a person with an FBS-r score of *T*≥80 is properly diagnosed with NES, regardless of the NUC score.

Sensitivity and specificity. It is notable that optimal classification alone does not give a full picture of differences seen at different levels of bias. As shown in Tables 11 and 12, sensitivity and specificity changed at different

levels of bias, even when classification rates were similar; for example, at F-r <55, similar classification rates of 69% were seen for RC1 ≥60 and RC1 ≥65; however, sensitivity dropped from 57% for RC1 ≥60 at to only 38% for RC1 ≥65, whereas specificity rose for those scores from 76% to 87%. Sensitivity tended to decrease as substantive scores increased, whereas specificity increased. In general, for F-r and Fs, specificity tended to be higher for low levels of bias and specificity tended to be higher for high levels of bias; the opposite tendency is seen for L-r and K-r. Interestingly, FBS-r tends to follow the same trend as F-r and Fs (that is, higher specificity at low levels of bias and higher sensitivity at higher levels of bias).

Positive Predictive Power and Negative Predictive Power. Similarly to sensitivity and specificity, PPP and NPP in this sample were not adequately represented by classification rates alone, as shown in Tables 13 and 14. In general, NPP tended to decrease as substantive scales increased, whereas PPP increased as substantive scores decreased. This was true regardless of level of bias indicator or whether the bias indicator represented NIM or PIM.

Discussion

The aim of this study was to examine the propensity of bias indicator scores to affect the ability of substantive indicators to distinguish between ES and NES, a conversion disorder that mimics of the physical manifestations of seizures without any underlying neurological activity. This study had two major goals: to psychometrically assess whether bias variables act in a moderating or suppressing manner in this situation and to determine the most accurate cut scores for clinicians working with ES and NES. Results

indicated that, in these data, bias indicators do indeed suppress substantive indicators, but moderation was not found. In response to this suppression, tables were created to improve accuracy of prediction of ES or NES based on substantive scales.

This present study has some methodological limitations regarding the use of overall classification rate to determine the best cut score. This method was chosen because it is intuitive for clinicians, as well as being the statistic chosen by Locke et al. (2010) in the study which the present study builds on. One alternative to classification rates is ROC curves, graphical plots illustrating sensitivity and specificity at all possible points on the variables. The advantage of ROC curves is that the *T*-scores can remain continuous, rather than looking at particular, previously determined cut scores; however, ROC curves, compared to overall classification rates, are less intuitive for clinician use. Additionally, the use of previously-determined cut scores was modeled after similar usage in the *Manual*, as well as by other researchers in the field.

Another alternative to classification rates for optimal cut score determination is examining the odds ratio, or the ratio of the odds of an event occurring in one group compared to the odds of it occurring in other group (e.g., being diagnosed with NES in a group with vEEG-diagnosed NES or vEEG-diagnosed ES). This method might be statistically preferable, since logistic regression, the statistic used in the present study, uses the odds ratio in its modeling. It was not used in this study for similar reasons as the ROC

curve; namely, it is not particularly clinician-friendly, and this study strives to be of optimal use clinically.

Another limitation of the study is the nature of the sample used. Well over 90% of the sample was Caucasian. Although this is a result of the true sample of inpatients at this particular location, it may not reflect psychosocial differences seen differentially among ethnic groups. Future studies should endeavor to ensure adequate representation among different ethnic groups.

Suppression but not moderation was found with these participants. Moderation, in which the relationship between the substantive variable and the outcome is different at different levels of bias, is a possibility when the two bias groups are fundamentally different, such as males and females (in some contexts) or people with entirely different medical conditions. But those participants with higher or lower levels of NIM or PIM are not fundamentally different from those with average symptom reporting, especially in the context of a disorder that has exaggerated symptoms as its hallmark. All persons exist on a continuum of response bias, either through unconscious reasons or accurate endorsement of problems (in the case of NIM), or good coping strategies (in the case of PIM). It therefore follows that suppression rather than moderation would be found in this sample.

These results add information to the bias indicator literature, and further dispute the findings of McGrath et al. (2010), which stated that bias indicators neither moderate nor suppress substantive indicators. Their conclusions were drawn in regard to several areas in which bias indicators

are used, including personality assessment and workplace variables. They further reported that dearth of data in other contexts, such as emotional disorders, forensic testing, and disability eligibility, indicate that utility of bias indicators is currently untested. The data from the present study indicate that in the context of psycho-medical testing, there are indeed suppression effects of MMPI-2-RF bias indicators on substantive indicators for ES and NES discrimination. The present author is unaware of other studies in this unique and very important context. These data therefore support the position of Rohling et al. (2011), who state that there is evidence in clinical situations (especially in clinical neuropsychology) to support the utility of bias indicators in enhancing prediction. At the least, these findings support the need for further studies on how bias can affect substantive scales.

Since suppression was found in the present study, additional clinically relevant statistics were run. Suppression was illustrated by changes in optimal cut scores for best overall classification; for negative impression management scales (i.e. F-r and Fs), optimal cut scores increased with higher bias indicator scores, whereas for positive impression management scales (i.e. L-r and K-r), optimal cut scores decreased with higher bias indicator scores. Although suppression was not seen for FBS-r, it was found that as FBS-r scores increased, the optimal cut score on NUC decreased. These changes in optimal cut scores indicate that it may be in a patient's best interest for a clinician, when determining the probability of a correct diagnosis, to choose take into account the patient's level of bias.

For the present participants, these changes in optimal cut scores are statistically significant, but it may be questioned whether they are clinically significant. Locke et al.'s 2010 study, using the same data, indicated that RC1 was the best predictor of NES or ES at a cut score of 65, with 68% accuracy. In the present study, the greatest increase in accuracy by looking at classification at different levels of bias was found at different levels of K-r; accuracy increased from 68% at a blanket RC1 *T*-score of 60 or 65 to 73% when calculated at optimal cut scores, an overall increase of 5% accuracy (this greatest increase in accuracy is consistent with the suppression analyses, in which K-r had the largest beta of the bias indicators)

However, the level of increase is not static across levels of bias. The optimal cut score for moderate K-r was indeed 65, but for the lowest level of K-r, a cut score of 65 only accurately predicts 62% of people, compared to 71% at RC1 = 75; and for high K-r, a cut score of 65 only accurately predicts 70% of people, compared to 76% at RC1 = 60. For moderate levels of K-r, the optimal cut score was RC1 = 65, so for people with moderate levels of K-r, there is no difference from the overall optimal cut score. In the sample of 131 people with low levels of K-r, an additional 12 people would be accurately diagnosed by using an optimal cut score as compared to a blanket score for all people (93 compared to 81); for the 146 people with high levels of K-r, 9 extra people are accurately diagnosed at a different cut score (111 vs. 102). It can be argued that these differences in overall hit rates are both statistically and clinically significant for people with high or low levels of K.

It is not clear why using K-r as a suppressor made the greatest difference of all the bias indicators. In an attempt to understand these results, the individual items were examined (see Table 15 for analyses). It was found that the K-r scale is composed of 10 "self" items (e.g. "I certainly feel useless at times") and four "other" items (e.g. "It takes a lot of argument to convince most people of the truth"). Individual analyses were then conducted of these 14 items. Five of the "self" items and all four "other" items were found to individually suppress RC1's diagnostic ability for ES/NES. Of the five "self" items, patients tended to respond in the opposite direction to good adjustment on three of those items, indicating that they felt useless and that they had many troubles piling on, and they frequently worried. The other two "self" items were about more social aspects of adjustment (i.e." I do not find it hard to make talk when I meet new people" and "I get mad easily easily and then get over it again soon"), and these were coded in the direction of greater adjustment. Of the four "other" items, NES patients coded other people as especially well-adjusted on all items. ES patients coded other as well-adjusted on two items (though not to the degree of the NES people) and not well-adjusted on two others. In summary, the K-r items that acted as suppressors were items in which patients (especially NES patients) indicate that other people are very well adjusted, but they themselves are not.

These analyses suggest several possible reasons why K-r worked best as a suppressor. The "other" items that were endorsed as false by patients, and also worked to discriminate between people with NES and ES, are all also coded on RC3, or cynicism. As stated in the *Manual*, a high RC1 and low

RC3 score is suspicious for a conversion disorder. It may be that people with lower levels of the RC3 items combined with higher "self" items may be underreporting somatization symptoms (as measured by RC1) to attempt to "fit in" with a world that they see as extraordinarily well-adjusted. Additionally, K-r is a measure of attitude, and questions about what others think or do can be seen as subtle measures of how the test-takers view themselves. That is, test-takers who state that others are well adjusted may be unconsciously suggesting that they themselves have those thoughts or tendencies. The more obvious "self" items that were coded as true endorse problems or anxieties that may directly relate to problems that they experience because of their illness, and is thus not inconsistent with a conceptualization of a generally emotionally healthy world.

Another important finding concerns the differences in sensitivity/specificity as the cut score changed. For the lowest levels of K-r, at the overall RC1 best cut score of T= 65, sensitivity was .87 and specificity was .43; at the optimal cut score for high K of RC1 T = 75, sensitivity was .67 and specificity was .74. Sensitivity decreased appreciably, whereas specificity increased appreciably. For clinicians choosing the cut score they wish to use, it may be important for them to choose not based on overall correct classification, but rather based on a careful examination of whether they are more interested in correctly classifying those with the disease (NES) or correctly classifying those with ES. If clinicians decide they would rather ensure that all patients with NES are flagged by the MMPI-2-RF (a relatively inexpensive and easy test to give as a first screener) and sent for further

testing at an EMU, they may choose the cut score of T= 65, even though the hit rate is lower due to the larger number of people inaccurately sorted as NES. On the other hand, if clinicians felt it was more important to avoid giving further expensive tests, such as an EMU stay, to people who do not need them, they may choose the higher cut score of T = 75, even though people with NES will be missed. Each clinician may wish to look at the trade-off between sensitivity and specificity at each cut score and choose the one that makes most sense for their patient base.

Likewise, if a clinician has a patient with a particular RC1 score, he or she can look at PPP/NPP to determine the likelihood that the patient was, in fact, accurately sorted as NES or ES. A patient who scored at RC1 T = 65 and had a low level of K-r bias would only be correctly identified as NES 53% of the time, or close to chance (PPP), but a person who scored lower than T = 65 would be correctly identified as ES 83% of the time (NPP); however, for a person with a high level of K-r bias, PPP is .78 and NPP is only .63. In other words, clinicians can be surer about whether they can confidently diagnose as ES or NES if they take bias into account.

The present study lends support to the practice of using bias indicators as suppressors and raises many interesting questions for future studies, while also providing tables for clinical use in discriminating between NES and ES for similar situations. It is hoped that future studies will investigate the use of suppressors in other psycho-medical settings, where bias scales may not be traditionally utilized beyond excluding persons that have bias scores above a particular cut score. Additionally, other tests such as the PAI may also

increase discrimination through accounting for bias suppression in NES/ES studies as well as other contexts. Although this study was able to determine the best suppressor for ES/NES discrimination in the present sample, it was not able to predict beforehand which scale would work best as a suppressor, as there is a dearth of literature as to what makes scales act as suppressors. This lack of research provides a further avenue of future studies. Pinpointing what makes a good suppressor will be a strong way to improve discrimination by narrowing the bias indicators to be tested, as well as lending itself to a greater understanding of what psychological forces can blur the usefulness of substantive scales.

References

Alper, K., Devinsky, O., Perrine, K., Vazquez, B., & Luciano, D. (1993). Nonepileptic seizures and childhood sexual and physical abuse. *Neurology, 43*, 1950.

American Psychiatric Association Task Force on DSM-IV (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR.* American Psychiatric Association.

Arnold, L. M., & Privitera, M. D. (1996). Psychopathology and trauma in epileptic and psychogenic seizure patients. *Psychosomatics, 37*, 438.

Babin, P. R., & Gross, P. (2002). Traumatic brain injury when symptoms don't add up: Conversion and malingering in the rehabilitation setting. *The Journal of Rehabilitation, 68*, 4-14.

Baer, R. A., Wetter, M. W., Nichols, D. S., Greene, R., & Berry, D. T. R. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment, 7*, 419.

Bagby, R. M., Marshall, M. B., Bury, A. S., Bacchiochi, J. R., & Miller, L. S. (2006). Assessing underreporting and overreporting response styles on the MMPI-2.

Bagby, R. M., Nicholson, R. A., Bacchiochi, J. R., Ryder, A. G., & Bury, A. S. (2002). The predictive capacity of the MMPI-2 and PAI validity scales and indexes to detect coached and uncoached feigning. *Journal of Personality Assessment, 78*, 69-86.

Bagby, R. M., Nicholson, R. A., Buis, T., & Bacchiochi, J. R. (2000). Can the MMPI-2 validity scales detect depression feigned by experts? *Assessment, 7*, 55.

Bagby, R. M., Rogers, R., Nicholson, R., Buis, T., Seeman, M. V., & Rector, N. (1997). Does clinical training facilitate feigning schizophrenia on the MMPI-2? *Psychological Assessment, 9*, 106.

Barr, W., Larson, E., Alpert, K., & Devinski, O. (2005). Rates of invalid MMPI-2 responding in patients with epileptic and nonepileptic seizures. *Epilepsia, 46.*

Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Mcnulty, J. L. (2002). The appropriateness of the MMPI-2 K correction. *Assessment, 9*, 219.

Benbadis, S. R. (2005a). The problem of psychogenic symptoms: Is the psychiatric community in denial? *Epilepsy & Behavior, 6*, 9-14.

Benbadis. S. R.. (2005b). Psychogenic non-epileptic seizures. In E. Wyllie (Ed.), *The treatment of epilepsy: Principles and practice. 4th edition (*pp. 623-630). Philadelphia, PA: Lippincott, Williams, & Wilkins.

Benbadis, S. R., & Allen Hauser, W. (2000). An estimate of the prevalence of psychogenic non-epileptic seizures. *Seizure, 9*, 280-281.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF: Manual for administration, scoring and interpretation.* University of Minnesota Press.

Betts, T., & Boden, S. (1992). Diagnosis, management and prognosis of a group of 128 patients with non-epileptic attack disorder. part II. previous childhood sexual abuse in the aetiology of these disorders. *Seizure, 1*, 27-32.

Bowman, E. S., & Markand, O. N. (1996). Psychodynamics and psychiatric diagnoses of pseudoseizure subjects. *American Journal of Psychiatry, 153*, 57.

Buchner, A., Erdfelder, E., Faul, F., & Lang, A. (2009). G*Power (Version 3.1.2)[Computer program]. Retrieved October 14, 2012, from http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/

Buderer, N. M. (1996). Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Academic Emergency Medicine, 3,* 895-900.

Burchett, D. L., & Ben-Porath, Y. S. (2010). The impact of overreporting on MMPI-2-RF substantive scale score validity. *Assessment, 17*, 497.

Butcher, J. N., Graham, J., Ben-Porath, Y., Tellegen, A., Dahlstrom, W., & Kaemmer, B. (2001). *MMPI-2: Minnesota multiphasic personality inventory-2: Manual for administration, scoring, and interpretation* University of Minnesota Press.

Cascino, G. D. (2002). Clinical indications and diagnostic yield of video-electroencephalographic monitoring in patients with seizures and spells. *Mayo Clinic Proceedings, 77*, 1111.

Cattell, R. B., Eber, H., & Tatsuoka, M. (1970). *Handbook for the 16PF.* Institute for Personality and Ability Testing, Illinois.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd Ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

Cragar, D. E., Berry, D. T. R., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2002). A review of diagnostic techniques in the differential diagnosis of epileptic and nonepileptic seizures. *Neuropsychology Review, 12*, 31-64.

Cragar, D. E., Berry, D. T. R., Schmitt, F. A., & Fakhoury, T. A. (2005). Cluster analysis of normal personality traits in patients with psychogenic nonepileptic seizures. *Epilepsy & Behavior, 6*, 593-600.

Cragar, D. E., Schmitt, F. A., Berry, D. T. R., Cibula, J. E., Dearth, C. M. S., & Fakhoury, T. A. (2003). A comparison of MMPI-2 decision rules in the diagnosis of nonepileptic seizures. *Journal of Clinical and Experimental Neuropsychology, 25*, 793-804.

D'Alessio, L., Giagante, B., Oddo, S., & Silva, W. (2006). Psychiatric disorders in patients with psychogenic non-epileptic seizures, with and without comorbid epilepsy. *Seizure, 15*, 333-339.

Derry, P. A., & McLachlan, R. S. (1996). The MMPI-2 as an adjunct to the diagnosis of pseudoseizures. *Seizure, 5*, 35-40.

Devinsky, O. (1996). Clinical profile of patients with epileptic and nonepileptic seizures. *Neurology, 46*, 1530.

Drake, M. E., Jr, Pakalnis, A., & Phillips, B. B. (1992). Neuropsychological and psychiatric correlates of intractable pseudoseizures. *Seizure : The Journal of the British Epilepsy Association, 1*, 11-13.

Eirís-Puñal, J., Rodríguez-Núñez, A., Fernández-Martínez, N., Fuster, M., Castro-Gago, M., & Martinón, J. (2001). Usefulness of the Head-Upright tilt test for distinguishing syncope and epilepsy in children. *Epilepsia, 42*, 709-713.

Ettinger, A. B., Devinsky, O., Weisbrot, D. M., Ramakrishna, R. K., & Goyal, A. (1999). A comprehensive profile of clinical, psychiatric, and psychosocial characteristics of patients with psychogenic nonepileptic seizures. *Epilepsia, 40*, 1292-1298.

Freud, S., & Breuer, J. (1895). *Studies on hysteria*. London: Hogarth Press

Gates, J. R. (2002). Nonepileptic seizures: Classification, coexistence with epilepsy, diagnosis, therapeutic approaches, and consensus. *Epilepsy and Behavior, 3*, 28-33.

Gervais, R., Lees-Haley, P., & Ben-Porath, Y. (2007). Predicting SVT performance with the MMPI-2-RF, FBS-r, RBS, and Fs scales. *Poster Presented at the 27th Annual Meeting of the National Academy of Neuropsychology, Scottsdale, AZ.*

Geyer, J. D., Payne, T. A., & Drury, I. (2000). The value of pelvic thrusting in the diagnosis of seizures and pseudoseizures. *Neurology, 54*, 227.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11*, 340-344.

Goldberg, J. F., & Burdick, K. E. (2001). Cognitive side effects of anticonvulsants. *Journal of Clinical Psychiatry, 62*, 27-33.

Gough, H. G. (1946). Diagnostic patterns on the Minnesota Multiphasic Personality Inventory. *Journal of Clinical Psychology.*

Hathaway, S., & McKinley, J. (1940). *The MMPI manual*. New York: Psychological Corporation.

Hsieh, F. Y., Block, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, *17*, 1623-1634.

Hurst, D. F., Locke, D. E. C., & Osborne, D. (2010). Evaluation of potential kidney donors with the personality assessment inventory: Normative data for a unique population. *Journal of Clinical Psychology in Medical Settings, 17*, 183-194.

Kuyk, J., Swinkels, W., & Spinhoven, P. (2003). Psychopathologies in patients with nonepileptic seizures with and without comorbid epilepsy: How different are they? *Epilepsy & Behavior, 4*, 13-18.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports.*

Lempert, T., Dieterich, M., Huppert, D., & Brandt, T. (1990). Psychogenic disorders in neurology: Frequency and clinical spectrum. *Acta Neurologica Scandinavica, 82*, 335-340.

Lesser, R. P., Lueders, H., & Dinner, D. S. (1983). Evidence for epilepsy is rare in patients with psychogenic seizures. *Neurology, 33*, 502.

Locke, D. E. C., Kirlin, K. A., Thomas, M. L., Osborne, D., Hurst, D. F., Drazkowski, J. F., & Noe, K. H. (2010). The Minnesota Multiphasic Personality Inventory-2-Restructured Form in the Epilepsy Monitoring Unit. *Epilepsy & Behavior, 17*, 252-258.

Locke, D. E., Kirlin, K. A., Wershba, R., Osborne, D., Drazkowski, J. F., Sirven, J. I., & Noe, K. H. (2011). Randomized comparison of the Personality Assessment Inventory and the Minnesota Multiphasic Personality Inventory-2 in the Epilepsy Monitoring Unit. *Epilepsy & Behavior : E&B,* doi:10.1016/j.yebeh.2011.05.023

Locke, D. E., & Thomas, M. L. (2011). Initial development of Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) scales to identify patients with psychogenic nonepileptic seizures. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 335-343. doi:10.1080/13803395.2010.518141

Lowry, R. (2001-2012). Significance of the difference between two correlation coefficients [software]. Retrieved October 14, 2012, from http://vassarstats.net/rdiff.html

Mari, F., Di Bonaventura, C., Vanacore, N., Fattouch, J., Vaudano, A. E., Egeo, G., & Giallonardo, A. T. (2006). Video-EEG study of psychogenic nonepileptic seizures: Differential characteristics in patients with and without epilepsy. *Epilepsia, 47*, 64-67.

Martin, R., Burneo, J., Prasad, A., Powell, T., Faught, E., Knowlton, R., & Kuzniecky, R. (2003). Frequency of epilepsy in patients with psychogenic seizures monitored by video-EEG. *Neurology, 61*, 1791.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*, 450-470.

McKinley, J., & Hathaway, S. (1944). The Minnesota Multiphasic Personality Inventory. V. hysteria, hypomania and psychopathic deviate. *Journal of Applied Psychology, 28*, 153-174.

Morey, L. C. (1991). *The personality assessment inventory* Psychological assessment resources.

Nelson, N., Sweet, J., & Demakis, G. (2006). Meta-analysis of the MMPI-2 fake bad scale: Utility in forensic practice. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Sec, 20*, 39-58.

Nelson, N. W., Hoelzle, J. B., Sweet, J. J., Arbisi, P. A., & Demakis, G. J. (2010). Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): Verified utility in forensic practice. *The Clinical Neuropsychologist, 24*, 701-724.

Nelson, N. W., Parsons, T. D., Grote, C. L., Smith, C. A., & Sisung, J. R. (2006). The MMPI-2 fake bad scale: Concordance and specificity of true and estimated scores. *Journal of Clinical and Experimental Neuropsychology, 28*, 1-12.

Paajanen, G., Hansen, T., & McLellan, R. (1993). PDI employment inventory and PDI customer service inventory manual. *Minneapolis, MN: Personnel Decisions, Inc.*

Reuber, M., & Elger, C. E. (2003). Psychogenic nonepileptic seizures: Review and update. *Epilepsy & Behavior, 4*, 205-216.

Reuber, M., Fernández, G., Helmstaedter, C., Qurishi, A., & Elger, C. E. (2002). Evidence of brain abnormality in patients with psychogenic nonepileptic seizures. *Epilepsy & Behavior, 3*, 249-254. doi:DOI: 10.1016/S1525-5050(02)00004-5

Reuber, M., Pukrop, R., Bauer, J., Derfuss, R., & Elger, C. E. (2004). Multidimensional assessment of personality in patients with psychogenic non-epileptic seizures. *Journal of Neurology, Neurosurgery & Psychiatry, 75*, 743-748. doi:10.1136/jnnp.2003.013821

Roelofs, K., Keijsers, G. P. J., Hoogduin, K. A. L., Naring, G. W. B., & Moene, F. C. (2002). Childhood abuse in patients with conversion disorder. *American Journal of Psychiatry, 159*, 1908.

Rogers, R. (2008). *Clinical assessment of malingering and deception.* The Guilford Press.

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders. *Assessment, 10*, 160.

Rohling, M. L., Larrabee, G. J., Greiffenstein, M. F., Ben-Porath, Y. S., Lees-Haley, P., Green, P., & Greve, K. W. (2011). A misleading review of response bias: Comment on McGrath, Mitchell, Kim, and Hough (2010).

Rosenberg, H. J., Rosenberg, S. D., Williamson, P. D., Wolford, I., & George, L. (2000). A comparative study of trauma and posttraumatic stress disorder prevalence in epilepsy patients and psychogenic nonepileptic seizure patients. *Epilepsia, 41*, 447-452.

Sar, V., Akyuz, G., Kundakci, T., Kiziltan, E., & Dogan, O. (2004). Childhood trauma, dissociation, and psychiatric comorbidity in patients with conversion disorder. *American Journal of Psychiatry, 161*, 2271.

Schachter, S. C., Brown, F., & James Rowan, A. (1996). Provocative testing for nonepileptic seizures: Attitudes and practices in the United States among American Epilepsy Society members. *Journal of Epilepsy, 9*, 249-252.

Sellbom, M., & Bagby, R. M. (2008). Validity of the MMPI-2-RF (Restructured Form) Lr and Kr scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20*, 370-376.

Sellbom, M., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., & Graham, J. R. (2006). Elevation differences between MMPI-2 clinical and restructured clinical (RC) scales. *Assessment, 13*, 430.

Sigurdardottir, K. R., & Olafsson, E. (1998). Incidence of psychogenic seizures in adults: A population-based study in Iceland. *Epilepsia, 39*, 749-752.

Smith, D., Defalla, B., & Chadwick, D. (1999). The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. *QJM, 92*, 15.

Storzbach, D., Binder, L., Salinsky, M., Campbell, B., & Mueller, R. (2000). Improved prediction of nonepileptic seizures with combined MMPI and EEG measures. *Epilepsia, 41*, 332-337.

Thompson, A. W., Hantke, N., Phatak, V., & Chaytor, N. (2010). The Personality Assessment Inventory as a tool for diagnosing psychogenic nonepileptic seizures. *Epilepsia, 51*, 161-164.

Wagner, M. T., Wymer, J. H., Topping, K. B., & Pritchard, P. B. (2005). Use of the personality assessment inventory as an efficacious and cost-effective diagnostic tool for nonepileptic seizures. *Epilepsy & Behavior, 7*, 301-304.

Warner, M., Wilkus, R., Vossler, D., Wyler, A., & Abson-Kraemer, D. (1996). MMPI-2 profiles in differential diagnosis of epilepsy vs. psychogenic seizures. *Epilepsia, 37*, 19.

Wilkus, R. J., Dodrill, C. B., & Thompson, P. M. (1984). Intensive EEG monitoring and psychological studies of patients with pseudoepileptic seizures. *Epilepsia, 25*, 100-107.

Wygant, D. B., Ben-Porath, Y. S., Arbisi, P. A., Berry, D. T. R., Freeman, D. B., & Heilbronner, R. L. (2009). Examination of the MMPI-2 restructured form (MMPI-2-RF) validity scales in civil forensic settings: Findings from simulation and known group samples. *Archives of Clinical Neuropsychology, 24*, 671.

Youngjohn, J. R., Wershba, R., Stevenson, M., Sturgeon, J., & Thomas, M. L. (2011). Independent validation of the MMPI-2-RF Somatic/Cognitive and validity scales in TBI litigants tested for effort. *The Clinical Neuropsychologist 25*, 463-76.

Table 1

Scales from MMPI-2-RF

Scale	Scale Name
Validity Scales	
?	Cannot Say
VRIN-r	Variable Response Inconsistency
TRIN-r	True Response Inconsistency
F-r	Infrequent Responses
Fp-r	Infrequent Psychopathology Responses
Fs	Infrequent Somatic Responses
FBS-r	Symptom Validity
L-r	Uncommon Virtues
K-r	Adjustment Validity
Restructured Clinical Scales	
RCd	Demoralization
RC1	Somatic Complaints
RC2	Low Positive Emotions
RC3	Cynicism
RC4	Antisocial Behavior
RC6	Ideas of Persecution
RC7	Dysfunctional Negative Emotion
RC8	Aberrant Experiences
RC9	Hypomanic Activation
Somatic/Cognitive Scales	
MLS	Malaise
GIC	Gastrointestinal Complaints
HPC	Head Pain Complaints
NUC	Neurological Complaints
COG	Cognitive Complaints

Table 2

Clinical Scales from MMPI-2

No.	Abbreviation	Scale Name
1	Hs	Hypochondriasis
2	D	Depression
3	Hy	Hysteria
4	Pd	Psychopathic Deviance
5	Mf	Masculinity-Femininity
6	Pa	Paranoia
7	Pt	Psychasthenia
8	Sc	Schizophrenia
9	Ma	Hypomania
0	Si	Social Introversion

Table 3

Decision Rules for Epileptic and Non-epileptic Seizures

Source	Decision Rule		
	1	2	3
Wilkus et al. (1984)	NES indicated if *any* of the following occur		
	Scale 1 or 3 is T ≥70 and is one of the two highest scales discounting Scales 5 or 0	Scale 1 or 3 is T ≥80	Scales 1 and 3 are T >59 and 10 points higher than Scale 2
Derry & McLachlan (1994)	NES indicated if *all* of the following occur		
	Scale 1 or 3 is T ≥70 and is one of the two highest scales discounting Scales 5 or 0	Scale 1 or 3 is T ≥80	Scales 1 and 3 are T >59 and 10 points higher than Scale 2
Modified Wilkus et al. (1984)	NES indicated if *any* of the following occur		
	Scale 1 or 3 is T ≥65 and is one of the two highest scales discounting Scales 5 or 0	Scale 1 or 3 is T ≥80	Scales 1 and 3 are T >59 and 10 points higher than Scale 2

Table 4

Demographics and History by Diagnostic Group

Variable	ES			NES			t/χ^2
	M	*SD*	*n*	*M*	*SD*	*n*	
Age	42.18	15.29	214	43.77	13.87	215	-1.12
Gender (female)	.64	—	214	.82	—	215	16.50**
Ethnicity (White)	.92	—	214	.94	—	215	.59
Handedness (right)	.86	—	214	.88	—	215	.24
Education	13.85	2.32	214	14.01	2.32	215	-.72
Frequency of seizures	1.92	1.92	214	1.44	1.44	215	3.63**
Current psychotropic medicines	.27	—	214	.53	—	215	30.12**
Presence of psychiatric history	.53	—	214	.81	—	215	34.97**
Presence of substance use history	.16	—	214	.19	—	215	.37
WRAT-4 Reading	99.93	9.32	194	99.48	11.67	182	.41
WAIS-III Full Scale IQ	99.73	11.91	191	100.60	13.84	158	-.63

Note. ES = Epileptic seizures; NES = Non-epileptic seizures. $^{**}p < .01$.

Table 5

Correlations between Phenotypic Variables

		1	2	3	4	5	6	7	8
1	ES/NES[a]	—							
2	RC1	.41**	—						
3	NUC	.40**	.81**	—					
4	F-r	.11*	.60**	.55**	—				
5	Fs	.15**	.64**	.59**	.60**	—			
6	L-r	.03	-.21**	-.22**	-.16**	-.15**	—		
7	K-r	.12*	-.30**	-.18**	-.55**	-.35**	.29**	—	
8	FBS-r	.32**	.79**	.57**	.49**	.53**	-.12*	-.18**	—

Note. [a]Criterion variable. ES = Epileptic seizures; NES = Non-epileptic seizures. *$p < .05$, **$p < .01$.

Table 6

Moderation Tests of the Interaction between Bias and Substantive Indicators

		RC1				NUC	
Bias indicator	Scales entered	*B*	*SE*	Bias indicator	Scales entered	*B*	*SE*
F-r	RC1	.095**	.012	F-r	NUC	.081**	.011
	F-r	-.029	.009		F-r	-.024*	.008
	RC1 x F-r	-.001	.000		NUC x F-r	.000	.001
F-s	RC1	.092**	.012	F-s	NUC	.081**	.011
	Fs	-.023**	.008		Fs	-.020**	.008
	RC1 x Fs	.000	.000		NUC x Fs	.001	.000
L-r	RC1	.077**	.010	L-r	NUC	.071**	.009
	L-r	.033**	.011		L-r	.031**	.011
	RC1 x L-r	.001	.001		NUC x L-r	.000	.001
K-r	RC1	.091**	.010	K-r	NUC	.075**	.009
	K-r	.069**	.013		K-r	.049**	.012
	RC1 x K-r	.001	.001		NUC x K-r	.001	.001
FBS-r	RC1	.072**	.014	FBS-r	NUC	.052**	.010
	FBS-r	.001	.012		FBS-r	.024**	.009
	RC1 x FBS-r	-.001	.001		NUC x FBS-r	.000	.001

Note. $^{*}p < .05$, $^{**}p < .01$.

Table 7

Correlations between Substantive Indicators and Diagnostic Outcome (Epileptic Seizures or Non-epileptic Seizures) at Different Levels of Bias Indicators[a]

	RC1	NUC	*n*
F-r <55	.39*	.32*	100
F-r =55-64	.42*	.36*	108
F-r =65-78	.49*	.50*	113
Fr ≥79	.35*	.37*	108
Fs <60	.32*	.37*	116
Fs =60-79	.43*	.41*	156
Fs ≥80	.39*	.41*	157
Fbs-r <55	.28*	.18*	99
Fbs-r =55-64	.40*	.41*	108
Fbs-r =65-79	.26*	.29*	124
Fbs-r ≥80	.26*	.25*	98
L-r <50	.38*	.43*	158
L-r =50-59	.41*	.42*	147
L-r ≥60	.48*	.42*	124
K-r <45	.43*	.37*	131
K-r =45-54	.44*	.41*	152
K-r ≥55	.50*	.50*	146

Note. [a] No correlations within each bias indicator were statistically different from each other. *$p < .05$.

Table 8

Regression Coefficients with the Addition of Bias Indicator for Each Pair of Substantive and Bias Indicators

		RC1				NUC	
Bias indicator	Scales entered	*B*	*SE*	Bias indicator	Scales entered	*B*	*SE*
No bias indicator	RC1	.069**	.009	No bias indicator	NUC	.065**	.008
F-r	RC1	.094**	.012	F-r	NUC	.081**	.011
	F-r	-.031**	.009		F-r	-.017**	.008
F-s	RC1	.091**	.012	F-s	NUC	.078**	.011
	Fs	-.024**	.008		Fs	-.017**	.007
L-r	RC1	.076**	.010	L-r	NUC	.071**	.009
	L-r	.031**	.011		L-r	.031**	.011
K-r	RC1	.089**	.010	K-r	NUC	.075**	.009
	K-r	.068**	.012		K-r	.048**	.012
FBS-r	RC1	.069**	.013	FBS-r	NUC	.052**	.010
	FBS-r	.000	.012		FBS-r	.024**	.009

Note. $^{**}p < .01$.

Table 9

True Positives (TP), False Positives (FP), False Negatives (FN), True Negatives (TR), Sensitivity (sens), Specificity (spec), and Classification Rates (CR) for F-r and RC1

		RC1 *T*-score								
		≥50	≥55	≥60	≥65	≥70	≥75	≥80	≥85	≥90
F-r <55	TP	34	27	21	14	7	2	1	1	0
	FP	42	26	15	8	4	0	0	0	0
	FN	3	10	16	23	30	35	36	36	37
	TN	21	37	48	55	59	63	63	63	63
	Sens	.92	.73	.57	.38	.19	.05	.03	.03	.00
	Spec	.33	.59	.76	.87	.94	1.00	1.00	1.00	1.00
	CR	.55	.64	**.69**	**.69**	.66	.65	.64	.64	.63
F-r =55-64	TP	54	51	49	38	27	10	6	3	1
	FP	48	41	25	16	9	3	1	0	0
	FN	1	4	6	17	28	45	49	52	54
	TN	5	12	28	37	44	50	52	53	53
	Sens	.98	.93	.89	.69	.49	.18	.11	.05	.02
	Spec	.09	.23	.53	.07	.83	.94	.98	1.00	1.00
	CR	.55	.58	**.71**	.69	.66	.56	.54	.52	.50
F-r =65-78	TP	61	61	61	60	53	35	24	13	7
	FP	51	48	41	30	20	10	7	3	2
	FN	0	0	0	1	8	26	37	48	54
	TN	1	4	11	22	32	42	45	49	50
	Sens	1.00	1.00	1.00	.98	.87	.57	.39	.21	.11
	Spec	.02	.08	.21	.42	.62	.81	.87	.94	.96
	CR	.55	.58	.64	.73	**.75**	.68	.61	.55	.50
F-r ≥79	TP	62	62	60	55	51	47	40	34	23
	FP	45	44	37	34	30	20	17	9	4
	FN	0	0	2	7	11	15	22	28	39
	TN	1	2	9	12	16	26	29	37	42
	Sens	1.00	1.00	.97	.89	.82	.76	.65	.55	.37
	Spec	.02	.04	.20	.26	.35	.57	.63	.80	.91
	CR	.58	.59	.64	.62	.62	**.68**	.64	.66	.60

Note. CRs in bold are the best classification rates for that bias level.

Table 10

Optimal Cut Scores and Classification Rates (CR) for All Levels of Bias [a]

	RC1 cut score	RC1 CR	NUC cut score	NUC CR
F-r <55	60 and 65	.69	85	.67
F-r =55-64	60	.71	75	.65
F-r =65-78	70	.75	85	.73
F-r ≥79	75	.68	90	.72
Fs <60	60	.70	70	.66
Fs =60-79	65 and 70	.71	80	.67
Fs ≥80	70 and 75	.68	85	.73
L-r <50	65 and 75	.66	85	.74
L-r =50-59	70	.70	85	.67
L-r ≥60	60	.76	70	.70
K-r <45	75	.71	90	.73
K-r =45-54	65 and 70	.72	80 and 85	.70
K-r ≥55	60	.76	70	.72
FBS-r <55	—	—	85	.76
FBS-r =55-64	—	—	80	.69
FBS-r =65-79	—	—	70	.62
FBS-r ≥80	—	—	≤55	.69

Note. [a] Cut scores and classification rates were not determined for different levels of FBS-r for RC1, since no suppression or moderation was found.

Table 11

Sensitivity (Sens) and Specificity (Spec) for RC1 at All Levels of Bias

	RC1 *T* ≥50		RC1 *T* ≥55		RC1 *T* ≥60		RC1 *T* ≥65		RC1 *T* ≥70		RC1 *T* ≥75		RC1 *T* ≥80		RC1 *T* ≥85		RC1 *T* ≥90	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
F-r <55	.92	.33	.73	.59	.57	.76	.38	.87	.19	.94	.05	1.00	.03	1.00	.03	1.00	.00	1.00
F-r =55-64	.98	.09	.93	.23	.89	.53	.69	.70	.49	.83	.18	.94	.11	.98	.05	1.00	.02	1.00
F-r =65-78	1.00	.02	1.00	.08	1.00	.21	.98	.42	.87	.62	.57	.81	.39	.87	.21	.94	.11	.96
F-r ≥79	1.00	.02	1.00	.04	.97	.20	.89	.26	.82	.35	.76	.57	.65	.63	.55	.80	.37	.91
Fs <60	.92	.32	.85	.44	.74	.67	.53	.79	.34	.86	.11	.97	.04	.98	.00	1.00	.00	1.00
Fs =60-79	1.00	.08	.91	.27	.90	.48	.75	.68	.54	.83	.28	.93	.16	.94	.07	.98	.06	1.00
Fs ≥80	1.00	.02	1.00	.05	.97	.19	.94	.25	.88	.38	.73	.60	.62	.07	.49	.84	.29	.90
L-r <50	.99	.07	.95	.15	.93	.28	.87	.46	.72	.57	.54	.78	.42	.83	.03	.94	.21	.96
L-r =50-59	.99	.09	.94	.20	.89	.47	.78	.59	.67	.73	.42	.85	.29	.91	.21	.95	.13	.97
L-r ≥60	.97	.26	.91	.49	.84	.67	.67	.77	.52	.86	.34	.93	.27	.93	.19	.95	.11	.98
K-r <45	1.00	.07	.98	.11	.87	.43	.87	.43	.82	.55	.67	.74	.53	.78	.44	.88	.27	.93
K-r = 45-54	.97	.09	.95	.20	.90	.42	.85	.59	.71	.73	.45	.86	.33	.89	.23	.96	.15	.99
K-r ≥55	.98	.25	.89	.50	.84	.66	.65	0.77	.46	.86	.27	.95	.02	1.00	.11	1.00	.05	1.00

Table 12

Sensitivity (Sens) and Specificity (Spec) for NUC at All Levels of Bias

	NUC *T* ≥50		NUC *T* ≥55		NUC *T* ≥60		NUC *T* ≥65		NUC *T* ≥70		NUC *T* ≥75		NUC *T* ≥80		NUC *T* ≥85		NUC *T* ≥90	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
F-r <55	1.00	.06	.95	.32	.65	.56	.65	.56	.51	.75	.30	.86	.16	.92	.11	1.00	.00	1.00
F-r =55-64	1.00	.00	.98	.06	.91	.23	.91	.23	.85	.42	.71	.58	.49	.77	.33	.92	.15	.96
F-r =65-78	1.00	.00	1.00	.04	.98	.12	.98	.12	.97	.23	.90	.29	.87	.50	.74	.73	.52	.88
F-r ≥79	1.00	.00	.98	.04	.98	.04	.98	.04	.95	.11	.89	.26	.85	.41	.81	.57	.71	.74
Fs <60	1.00	.06	.96	.33	.72	.52	.72	.52	.58	.73	.40	.81	.38	.87	.26	.94	.08	.95
Fs =60-79	1.00	.00	.99	.06	.94	.23	.94	.23	.91	.38	.79	.52	.59	.73	.44	.85	.26	.95
Fs ≥80	1.00	.00	.99	.02	.99	.03	.99	.03	.97	.11	.90	.24	.84	.40	.78	.67	.66	.79
L-r <50	1.00	.01	.99	.10	.95	.14	.95	.15	.91	.23	.84	.35	.79	.57	.71	.77	.57	.85
L-r =50-59	1.00	.01	.99	.09	.92	.25	.92	.25	.88	.43	.72	.60	.64	.68	.54	.80	.38	.92
L-r ≥60	1.00	.04	.97	.21	.85	.42	.85	.42	.78	.61	.66	.67	.49	.81	.36	.93	.21	.96
K-r <45	1.00	.00	.98	.05	.96	.16	.96	.16	.91	.29	.78	.38	.71	.54	.65	.70	.55	.86
K-r = 45-54	1.00	.03	.96	.12	.88	.22	.88	.22	.85	.36	.78	.53	.72	.69	.56	.85	.44	.89
K-r ≥55	1.00	.03	1.00	.22	.89	.42	.89	.42	.83	.58	.68	.69	.54	.81	.45	.94	.24	.98
FBS-r <55	1.00	.03	.89	.21	.57	.41	.57	.41	.43	.65	.36	.76	.29	.87	.29	.94	.07	.99
FBS-r =55-64	1.00	.02	1.00	.15	.94	.31	.94	.31	.83	.48	.64	.64	.51	.82	.36	.89	.26	.95
FBS-r =64-79	1.00	.02	.99	.06	.96	.13	.96	.13	.93	.19	.81	.29	.67	.46	.57	.75	.39	.88
FBS-r ≥80	1.00	.00	1.00	.00	.97	.00	.97	.00	.97	.03	.91	.13	.87	.27	.75	.53	.62	.67

Table 13

Positive Predictive Power (PPP) and Negative Predictive Power (NPP) for RC1 for All Variables

	RC1 *T* ≥50		RC1 *T* ≥55		RC1 *T* ≥60		RC1 *T* ≥65		RC1 *T* ≥70		RC1 *T* ≥75		RC1 *T* ≥80		RC1 *T* ≥85		RC1 *T* ≥90	
	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP
F-r <55	.45	.88	.51	.79	.58	.75	.64	.71	.64	.66	1.00	.64	1.00	.64	1.00	.64	—	.63
F-r =55-64	.53	.83	.55	.75	.66	.82	.70	.69	.75	.61	.77	.53	.86	.51	1.00	.50	1.00	.50
F-r =65-78	.54	1.00	.56	1.00	.60	1.00	.67	.96	.73	.80	.78	.62	.77	.55	.81	.51	.78	.48
F-r ≥79	.58	1.00	.58	1.00	.62	.82	.62	.63	.63	.59	.70	.63	.70	.57	.79	.57	.85	.52
Fs <60	.53	.83	.56	.78	.65	.75	.68	.67	.67	.61	.75	.56	.67	.55	—	.54	—	.54
Fs =60-79	.46	1.00	.49	.80	.57	.86	.65	.78	.71	.70	.76	.63	.69	.59	.71	.58	1.00	.58
Fs ≥80	.60	1.00	.61	1	.64	.80	.65	.73	.68	.69	.73	.60	.75	.55	.82	.52	.82	.46
L-r <50	.50	.86	.51	.75	.55	.82	.60	.79	.61	.69	.69	.64	.70	.61	.82	.59	.84	.57
L-r =50-59	.51	.88	.53	.79	.62	.81	.64	.73	.71	.70	.73	.60	.75	.57	.79	.55	.82	.54
L-r ≥60	.61	.88	.68	.82	.75	.78	.78	.67	.81	.60	.85	.55	.82	.52	.81	.50	.86	.52
K-r <45	.44	1.00	.44	.89	.53	.83	.53	.83	.57	.81	.65	.76	.63	.69	.73	.68	.75	.64
K-r = 45-54	.53	.78	.56	.79	.62	.79	.69	.79	.73	.70	.78	.60	.76	.56	.86	.54	.92	.53
K-r ≥55	.63	.89	.70	.78	.76	.76	.78	.63	.81	.56	.88	.50	1	.49	1.00	.47	1.00	.45

Note. Dashes indicate that PPP or NPP could not be calculated because of a zero in the denominator.

Table 14

Positive Predictive Power (PPP) and Negative Predictive Power (NPP) for NUC for All Variables

	NUC *T* ≥50		NUC *T* ≥55		NUC *T* ≥60		NUC *T* ≥65		NUC *T* ≥70		NUC *T* ≥75		NUC *T* ≥80		NUC *T* ≥85		NUC *T* ≥90	
	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP	PPP	NPP
F-r <55	.39	1.00	.45	.91	.46	.73	.46	.73	.54	.72	.55	.68	.55	.65	1.00	.66	—	.63
F-r =55-64	.51	—	.52	.75	.55	.71	.55	.71	.60	.73	.64	.66	.69	.59	.82	.57	.80	.52
F-r =65-78	.54	—	.55	1.00	.57	.86	.57	.86	.60	.86	.60	.71	.67	.76	.76	.70	.84	.61
F-r ≥79	.57	—	.58	.67	.58	.67	.58	.67	.59	.63	.62	.63	.66	.68	.71	.68	.79	.65
Fs <60	.47	1.00	.55	.91	.56	.69	.56	.69	.65	.68	.64	.61	.71	.63	.78	.60	.57	.55
Fs =60-79	.44	—	.45	.83	.48	.83	.48	.83	.53	.85	.56	.77	.63	.70	.70	.66	.82	.63
Fs ≥80	.60	—	.60	.50	.60	.67	.60	.67	.62	.70	.64	.63	.68	.63	.78	.67	.83	.61
L-r <50	.48	1.00	.50	.89	.50	.75	.51	.75	.52	.73	.55	.71	.63	.75	.74	.74	.78	.68
L-r =50-59	.49	1.00	.51	.88	.54	.76	.54	.76	.59	.78	.63	.69	.66	.66	.72	.65	.82	.61
L-r ≥60	.55	1.00	.59	.86	.63	.71	.63	.71	.70	.70	.70	.62	.75	.58	.86	.55	.88	.51
K-r <45	.42	—	.43	.80	.45	.86	.45	.86	.48	.81	.48	.71	.53	.72	.61	.74	.73	.72
K-r = 45-54	.52	1.00	.54	.75	.54	.64	.54	.64	.58	.69	.64	.70	.71	.70	.80	.65	.81	.60
K-r ≥55	.57	1.00	.62	1	.66	.75	.66	.75	.72	.73	.74	.63	.79	.58	.90	.57	.95	.50
FBS-r <55	.29	1.00	.31	.83	.28	.71	.28	.71	.32	.74	.37	.75	.47	.76	.67	.77	.67	.73
FBS-r =55-64	.44	1.00	.47	1.00	.51	.86	.51	.86	.55	.78	.58	.70	.69	.68	.71	.64	.80	.62
FBS-r =64-79	.59	1.00	.59	.75	.61	.70	.61	.70	.61	.67	.61	.52	.63	.50	.76	.56	.82	.51
FBS-r ≥80	.69	—	.69	—	.69	.00	.69	.00	.69	.33	.70	.40	.73	.47	.78	.48	.81	.43

Note. Dashes indicate that PPP or NPP could not be calculated because of a zero in the denominator.

Table 15

Individual K Items' Ability to Differentiate between Epileptic Seizures and Non-epileptic Seizures

K item	Item description	Coded "true" overall	Coded "true" ES only	Coded "true" NES only
23	At times I feel like smashing things.	.28	—	—
72	At times my thoughts have raced ahead faster than I could speak them.	.75[d]	.67	.82
202	I have never felt better in my life than I do now.	.08[c]	—	—
322	Criticism or scolding hurts me terribly.	.47[c]	—	—
338	I frequently worry	.43	—	—
10[a]	It takes a lot of argument to convince most people of the truth[b]	.31[d]	.36	.26
36[a]	I think a great many people exaggerate their misfortunes in order to gain the sympathy and help of others[b]	.47[d]	.58	.36
44[a]	I find it hard to make talk with others	.35[d]	.40	.30
89[a]	I certainly feel useless at times	.54	—	—
99[a]	Most people will use somewhat unfair means to gain profit or an advantage rather than to lose it[b]	.41[d]	.47	.35
155[a]	I get mad easily and then get over it soon	.36[d]	.41	.31
171[a]	I think nearly anyone would tell a lie to keep out of trouble[b]	.47[d]	.55	.40
187[a]	I have sometimes felt that difficulties were piling up so high that I couldn't overcome them	.58	—	—
338[a]	I frequently find myself worrying about something	.60	—	—

Note. [a] Denotes items that suppress RC1's ability to differentiate ES and NES. [b]Denotes items with an "other" rather than "self" focus. [c]Denotes items coded such that "true" is indicative of defensiveness [d]Denotes differences in % coded true for ES and NES at $p < .05$.

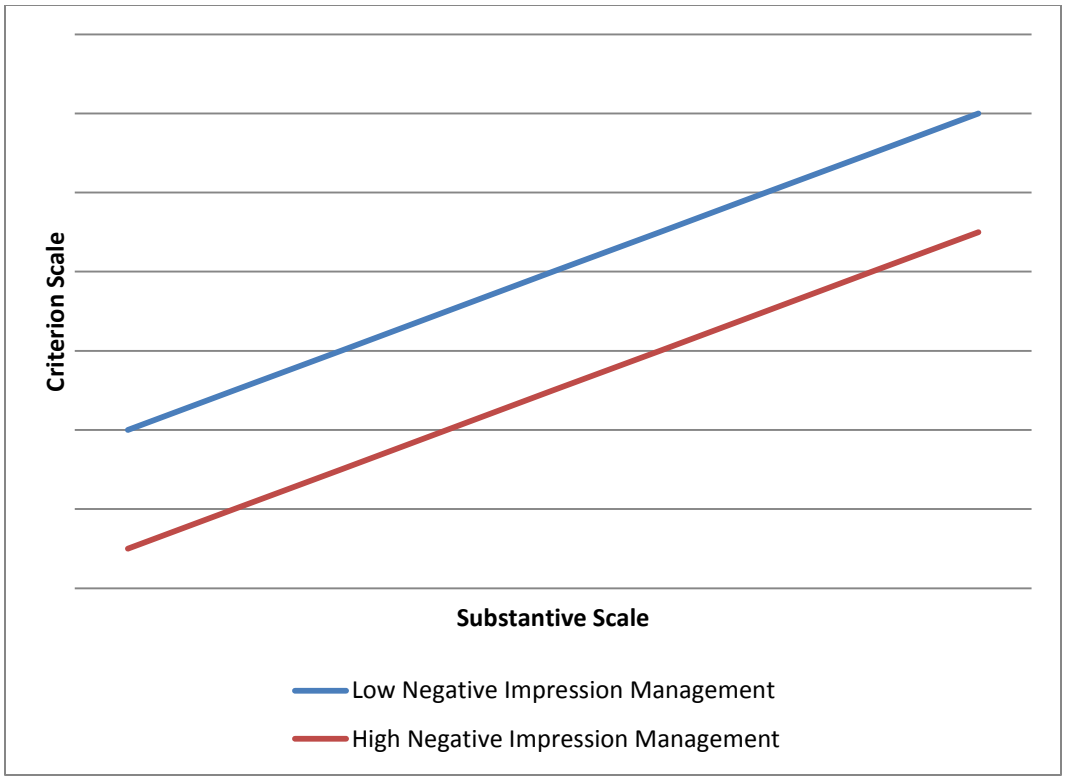

Figure 1. Illustration of suppression. In this example, overresponders have a similar regression slope to normal responders, but the same substantive scale score consistently predicts a lower criterion score.

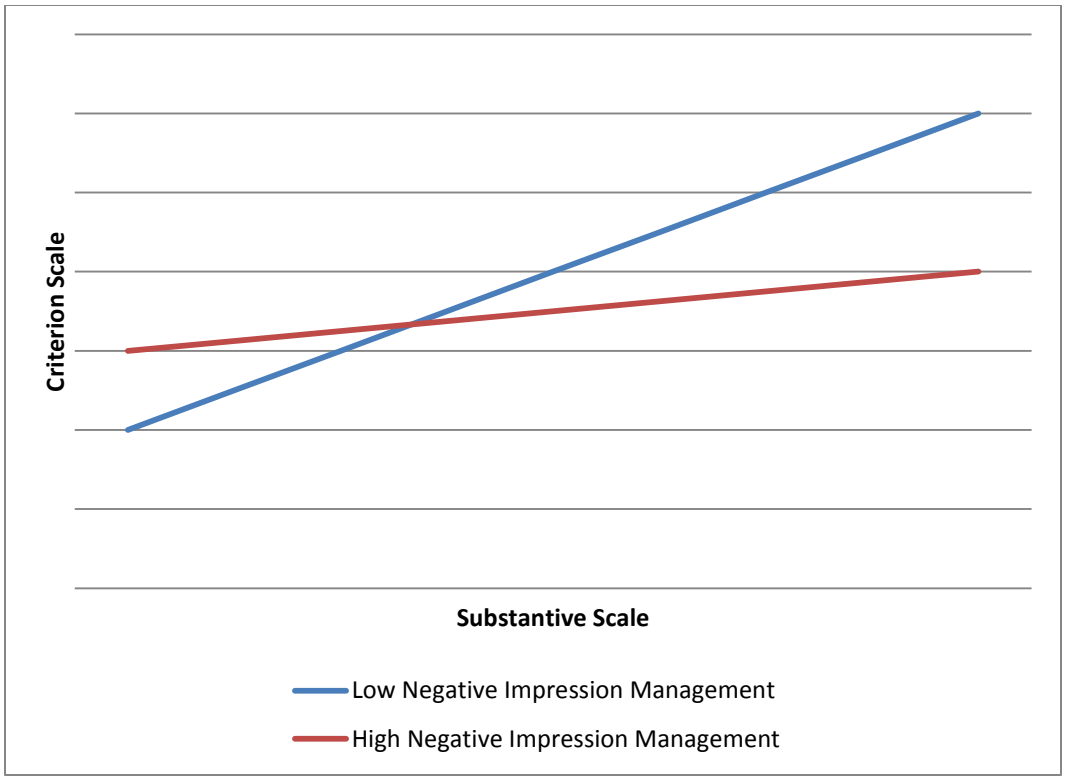

Figure 2. Illustration of moderation. In this example, the substantive scale cannot be used to predict the criterion with any degree of accuracy for those with high negative impression management.