MMPI-2-RF VALIDITY SCALE SCORES AS MODERATORS OF SUBSTANTIVE SCALE CRITERION VALIDITY

A dissertation submitted to Kent State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

by

Danielle L. Burchett

December, 2012

Dissertation written by Danielle Burchett B.S., California Polytechnic State University, San Luis Obispo, 2002 M.A., Kent State University, 2009 Ph.D., Kent State University, 2012

Approved by

Yossef S. Ben-Porath, Ph.D., Chair, Doctoral Dissertation Committee

John Gunstad, Ph.D., Members, Doctoral Dissertation Committee

Manfred van Dulmen, Ph.D.,

Daniel J. Flannery, Ph.D.,

Richard Meindl, Ph.D.,

Accepted by

| Maria Zaragoza, Ph.D., | Chair, Department of Psychology |
|---------------------------|------------------------------------|
| John R.D. Stalvey, Ph.D., | Dean, College of Arts and Sciences |

Table of Contents

| List of Figures | Х |
|---|----|
| List of Tables | xi |
| Acknowledgements | xv |
| I. Introduction | 1 |
| What Is Personality Assessment? | 2 |
| Types of Invalid Responding | 3 |
| Intentional vs. Unintentional Invalid Responding | 4 |
| Misleading Responding Defined | 6 |
| Models of Misleading Responding | 6 |
| Pathogenic Model of Misleading Responding | 7 |
| Criminological Model of Misleading Responding | 8 |
| Adaptational Model of Misleading Responding | 9 |
| Interpersonal Management Model | 10 |
| Interactional Model of Applicant Faking | 11 |
| Interactionist Model of Item-Level Response Distortion | 13 |
| Factors Associated with Misleading Responding | 14 |
| Setting/Context Factors Associated with Misleading Responding | 15 |
| Mental Health Settings | 15 |

| Emergency Departments and Substance Treatment Centers | 16 |
|---|----|
| Forensic Settings | 17 |
| Employment Settings | 19 |
| Individual Factors Associated with Misleading Responding | 20 |
| Sex | 20 |
| Race | 21 |
| Age | 23 |
| Educational and Vocational Attainment | 25 |
| Intellectual Abilities and Mental Retardation | 26 |
| Amnesia and Traumatic Brain Injury | 27 |
| Dementia | 28 |
| Depression | 29 |
| Anxiety | 29 |
| Trauma and Posttraumatic Stress Disorder | 30 |
| Schizophrenia and Other Psychotic Disorders | 33 |
| Eating Pathology and Obesity | 34 |
| Pathological Gambling | 35 |
| Violent Behaviors | 35 |
| Personality Dysfunction | 36 |
| Coaching and Practice | 36 |
| Strategies for Detecting Misleading Responding in Personality | |

| Assessment | 37 |
|--|----|
| Detecting Overreported Psychopathology | 37 |
| Detecting Underreported Psychopathology | 39 |
| Measures for Detecting Misleading Responding | 40 |
| The MMPI Family | 41 |
| Personality Assessment Inventory | 44 |
| Millon Clinical Multiaxial Inventory – III | 45 |
| Structured Interview of Reported Symptoms | 46 |
| Structured Inventory of Malingered Symptomatology | 47 |
| Miller Forensic Assessment of Symptoms Test | 47 |
| Misleading Responding Research Methodology | 49 |
| Application of Test Bias Identification Tools to Misleading Responding | |
| Research | 52 |
| Measuring External Test Bias | 52 |
| Using a Series of Multiple Linear Regressions | 54 |
| Using Hierarchical Moderated Multiple Linear Regressions | 55 |
| Using Hierarchical Moderated Multiple Logistic Regressions | 55 |
| Using External Test Bias Detection Strategies to Study Misleading | |
| Responding Indicators | 56 |
| Statistical Significance | 57 |
| Practical Significance | 58 |

| Previous Research on Validity Indices and Moderators of Criterion | |
|--|----|
| Validity | 61 |
| Non-Content-Based Invalid Responding Validity Scales as Moderators | 61 |
| Underreporting Validity Scales as Moderators | 64 |
| Overreporting Validity Scales as Moderators | 66 |
| Limitations of Existing Validity Scale Moderation Literature | 67 |
| The Current Study | 69 |
| Research Questions | 69 |
| II. Method | 73 |
| Participants | 73 |
| College Student Controls and Overreporting Simulators | 73 |
| Community Mental Health Center Outpatients | 74 |
| Psychiatric Inpatients | 75 |
| Criminal Forensic Assessment Center Defendants | 75 |
| Instruments and Measures | 76 |
| Minnesota Multiphasic Personality Inventory—2 Restructured Form | |
| (MMPI-2-RF) | 76 |
| Antisocial Behavior Questionnaire (ABQ) | 81 |
| Alcohol Use Disorders Identification Test (AUDIT) | 81 |
| Cognitive Difficulties Scale – 26-Item Version (CDS-26) | 82 |
| Drug Abuse Screening Test – 20-Item Version (DAST-20) | 82 |

| Emotionality, Activity, Sociability, Impulsivity Temperament Survey | |
|---|----|
| – II (EASI-II) | 82 |
| Goldberg Mania Scale (GMS) | 84 |
| General Self-Efficacy Scale (GSE) | 85 |
| Rosenberg Self-Esteem Scale (SES) | 85 |
| Symptom Checklist 90 – Revised (SCL-90-R) | 85 |
| Forensic Record Review Form (FRRF) | 85 |
| Inpatient Record Review Form (IRRF) | 86 |
| Patient Description Form (PDF) | 87 |
| Procedures | 87 |
| Operationalization of Study Hypotheses | 89 |
| Research Question One | 89 |
| Operationalization of Research Question One | 90 |
| Research Question Two | 91 |
| Operationalization of Research Question Two | 92 |
| III. Results | 93 |
| Bivariate Correlations | 93 |
| College Sample | 94 |
| Outpatient Sample | 95 |
| Inpatient Sample | 96 |
| Forensic Sample | 97 |

| Moderated Linear and Logistic Regression Analyses | 122 |
|--|-----|
| College Sample | 122 |
| Outpatient Sample | 124 |
| Inpatient Sample | 125 |
| Forensic Sample | 126 |
| Chi Square Summary Analyses | 152 |
| Differences in Overall Moderation Effects by Sample | 152 |
| Differences in Overall Moderation Effects by Validity Scale | 153 |
| Follow-Up Analyses | 153 |
| Independence of the Moderators (Validity Scales) and Dependent | |
| Variables (Criterion Measures) | 159 |
| Multicollinearity Between the Independent Variables (RC Scales) and | |
| Moderator Variables (Validity Scales) | 170 |
| IV. Discussion | 173 |
| Validity Scales as Moderators of Substantive Scale Criterion Validity | 174 |
| Moderation Due to Slope and Intercept Differences | 175 |
| The Trade-Off Between Generalizability and Statistical Accuracy | 177 |
| Support for Each Individual Validity Scale as a Moderator of Criterion | |
| Validity | 178 |
| Considerations from the Follow-Up Analyses | 179 |
| Implication of the Study Findings | 180 |

| Clinical Implications | 182 |
|--|-----|
| Limitations & Future Directions | 183 |
| Appendix A: Example SPSS Syntax for Running Hierarchical Moderated | |
| Linear and Logistic Regressions | 189 |
| References | 191 |

List of Figures

| Figure 1: Interactional Model of Applicant Faking (Snell, Sydell, & Lueke, | |
|---|-----|
| 1999) | 12 |
| Figure 2: Tett et al.'s (2006) Interactionist Model of Item-Level Response | |
| Distortion: Complete Lists of the Various Situational Factors, Abilities, and | |
| Nontargeted Traits | 14 |
| Figure 3: Types of External Test Bias | 54 |
| Figure 4: Example of Using Moderated Regression Analyses to Examine the | |
| Utility of the Validity Scales to Moderate Criterion Validity of Substantive | |
| Scales | 60 |
| Figure 5: Example of How Intercept Differences Lead to Differential | |
| Prediction of External Criterion Scores | 60 |
| Figure 6: Hypothesized and Other Graphical Depictions of Slope and | |
| Intercept Differences | 72 |
| Figure 7: Interpreting Significant Intercept Differences in the Unanticipated | |
| Direction (in the Presence of Anticipated Significant Slope | |
| Differences) | 176 |

List of Tables

| Table 1: Types of Personality Testing Invalid Responding | 5 |
|---|-----|
| Table 2: The MMPI-2-RF Scales | 77 |
| Table 3: Instruments and Measures Used in the Present Study | 83 |
| Table 4: College Sample Bivariate Correlations Between MMPI-2-RF | |
| Restructured Clinical Scales and Conceptually-Relevant External | |
| Criteria | 99 |
| Table 5: Outpatient Sample Bivariate Correlations Between MMPI-2- | |
| RF Restructured Clinical Scales and Conceptually-Relevant External | |
| Criteria | 102 |
| Table 6: Inpatient Sample Bivariate Correlations Between MMPI-2- | |
| RF Restructured Clinical Scales and Conceptually-Relevant External | |
| Criteria | 110 |
| Table 7: Forensic Sample Bivariate Correlations Between MMPI-2-RF | |
| Restructured Clinical Scales and Conceptually-Relevant External | |
| Criteria | 118 |
| Table 8: Significance Levels for College Sample Moderated Linear | |
| Regressions, Slope Differences, and Intercept Differences, and | |
| Associated R ² Values (<i>N</i> 's Range From 220 to 221) | 128 |

| Table 9: Significance Levels for Outpatient Sample Moderated Linear | |
|---|-----|
| Regressions, Slope Differences, and Intercept Differences, and | |
| Associated R ² Values (<i>N</i> 's Range From 707 to 748) | 134 |
| Table 10: Significance Levels for Inpatient Sample Moderated | |
| Logistic Regressions, Slope Differences, and Intercept Differences, | |
| and Associated Nagelkerke Pseudo-R ² Values (N's Range From 2,369 | |
| to 2,544) | 143 |
| Table 11: Significance Levels for Forensic Sample Moderated Linear | |
| & Logistic Regressions, Slope Differences, and Intercept Differences, | |
| and Associated R^2 & Nagelkerke Pseudo- R^2 Values (<i>N</i> 's Range From | |
| 1,138 to 1,358) | 148 |
| Table 12: Chi Square Analyses Examining Frequency of Statistically | |
| Significant Overall Moderation Effects by Sample | 154 |
| Table 13: Chi Square Analyses Examining Frequency of Statistically | |
| Significant Overall Moderation Effects by Scale Within the College | |
| Sample | 155 |
| Table 14: Chi Square Analyses Examining Frequency of Statistically | |
| Significant Overall Moderation Effects by Scale Within the Outpatient | |
| Sample | 156 |
| Table 15: Chi Square Analyses Examining Frequency of Statistically | |
| Significant Overall Moderation Effects by Scale Within the Inpatient | |

| Sample | 157 |
|--|-----|
| Table 16: Chi Square Analyses Examining Frequency of Statistically | |
| Significant Overall Moderation Effects by Scale Within the Forensic | |
| Sample | 158 |
| Table 17: Bivariate Correlations Between Validity Scale Moderators | |
| and (a) Independent Variable Restructured Clinical (RC) Scales and | |
| (b) Dependent Variable External Criterion Measures in the College | |
| Sample (<i>N</i> 's Range from 220 to 221) | 161 |
| Table 18: Bivariate Correlations Between Validity Scale Moderators | |
| and (a) Independent Variable Restructured Clinical (RC) Scales and | |
| (b) Dependent Variable External Criterion Measures in the Outpatient | |
| Sample (<i>N</i> 's Range from 707 to 748) | 163 |
| Table 19: Point Biserial Correlations Between Validity Scale | |
| Moderators and (a) Independent Variable Restructured Clinical (RC) | |
| Scales and (b) Dependent Variable External Criterion Measures in the | |
| Inpatient Sample (<i>N</i> 's Range from 2,369 to 2,544) | 166 |
| Table 20: Bivariate and Point Biserial Correlations Between Validity | |
| Scale Moderators and (a) Independent Variable Restructured Clinical | |
| (RC) Scales and (b) Dependent Variable External Criterion Measures | |
| in the Forensic Sample (<i>N</i> 's Range from 1,138 to 1,358) | 169 |
| Table 21: Recommendations for Best Practice in Examining Whether | |

Validity Scales Moderate Substantive Scale Criterion Validity...... 185

Acknowledgements

Reflecting on what got me to this point, I cannot help but appreciate all of the support from those who made this possible. Without the guidance from C.H., I would have never developed the drive to make it this far. Without J.B. and S.S., I would have never developed the backbone to put up with difficult days. Without the unconditional support and praise from F.F. and M.F., I would have difficulty appreciating how much my journey has meant to my family. And, without the eternal patience and sacrifice of J.S., it would have been impossible to balance my education and personal life. To all of my family: thank you so very much for your support. Words can't express my deep appreciation! I also wish to thank those I've met throughout my higher education who guided me down my current path. Without my undergraduate advisor, Dr. Donald Ryujin, I would have never learned to love psychology in the first place. Without my graduate advisor, Dr. Yossi Ben-Porath, I would have never learned so much about assessment, articulately explaining myself, and diligence. I also greatly appreciate what I've learned from Dr. Jack Graham, Dr. John Gunstad, and Dr. Manfred van Dulmen, whose lessons will be integral to the rest of my career. Finally, I wish to acknowledge my appreciation for the mental health clients who have enriched my life and helped me to put into perspective what is really most important: "The foundation of greatness is honoring the small things of the present moment, instead of pursuing the idea of greatness." -Eckhart Tolle

MMPI-2-RF Validity Scale Scores as Moderators of Substantive Scale Criterion Validity

Psychological assessment is a useful endeavor that provides a plethora of clinical information to clinicians, teachers, and employers, among other professionals. However, a variety of factors lead assessment to be an imperfect science. In addition to psychometric limitations of assessment instruments, the utility of test results is also impacted by the validity of responses given by test takers. By its nature, invalid responding is difficult to detect and measure. However, some researchers have attempted to estimate its prevalence across evaluation settings.

Mittenberg, Patton, Canyock, and Condit (2002) surveyed American Board of Clinical Neuropsychology (ABCN) diplomates, who estimated that about 29% of personal injury cases, 30% of disability and worker's compensation litigation cases, 19% of criminal cases, and 8% of general medical cases involve probable malingering and symptom exaggeration. In other studies, clinicians have estimated that malingering occurs in about 15 to 18% of forensic cases (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998; Rogers, Sewell, & Goldstein, 1994). In a study examining underreporting of problems in job applicant testing, managers believed that over half of their applicants engaged in some form of "faking behavior," (Rees & Metcalfe, 2003).

Regardless of the exact percentages, inaccurate responding is a significant problem in psychological testing. Examiners rely on test results to make inferences that may affect psychological and medical treatment, legal outcomes, and hiring decisions of evaluated persons. Although invalid responding occurs in other types of assessment (i.e., neuropsychological or medical evaluations), the focus of this study is on personality assessment. This chapter begins with a definition of personality assessment, followed by a review of the various types of invalid or misleading responding that can occur. Next, the existing models of misleading responding are reviewed, and specific setting and individual factors associated with it are identified. Strategies and validity scales used for the detection of misleading responding are discussed next, followed by research designs used to study this phenomenon and the findings of studies of misleading responding across a variety of settings. A discussion of how test bias detection methods (i.e., slope bias & intercept bias) can be applied to the area of misleading responding follows, leading to the research questions and hypotheses associated with the current investigation.

What is Personality Assessment?

Personality assessment, in its broadest sense, can be defined as the measurement of individual characteristics such as interests, attitudes, psychological traits and states, cognitive style, and self-identity (Cohen & Swerdlik, 2010). Personality assessment tools come in a wide variety of forms, including performance-based instruments and self-report inventories. Performance-based measures typically require the respondent to explain what they see or think in relation to ambiguous stimuli. Although these have been important tools in personality assessment for many decades, they are beyond the scope of this paper and will not be discussed further. Self-report inventories require the test-taker to respond to statements (items), typically by indicating whether or to what extent they relate to their own functioning. Various types of invalid responding to self-report items are described next.

Types of Invalid Responding

Ben-Porath (2003; Table 1) discussed the threats to protocol validity in personality assessment. These threats are typically classified as either *non-content-based* or *content-based* invalid responding. Non-content-based invalid responding occurs when an individual's invalid test responses are not related to the content of test items. This is further broken into *non- responding*, *random responding*, and *fixed responding*. Nonresponding occurs when the individual does not respond to test items. Random responding occurs when the individual indiscriminately responds to items with varying answers. Fixed responding occurs when the individual indiscriminately provides the same response to test items.

With content-based invalid responding, the individual pays attention to the content of the items, but provides responses that depict a distorted picture of their actual functioning. This includes *overreporting* and *underreporting* of symptoms. Overreporting occurs when individuals' responses lead them to appear worse off than they actually are. This can be intentional or unintentional. Test-takers sometimes intentionally *exaggerate* their current problems or *fabricate* problems they do not have. One example of fabrication or exaggeration, *malingering*, is by definition externally-motivated, context-

specific behavior that involves intentionally overreporting symptoms (Drob, Meehan, & Waxman, 2009; American Psychiatric Association, 2000). According to the DSM-IV-TR, malingering should be strongly suspected if any combination of the following is present: a medicolegal context for the evaluation, a significant discrepancy between subjective complaints and objective findings, a lack of cooperation in evaluation or treatment, or the presence of antisocial personality disorder. It is important to distinguish between malingering and Factitious Disorder, in which secondary gain incentives are absent, and Conversion Disorder, in which somatic symptoms are not intentionally produced (DSM-IV-TR; 2000). It is also important to note that the DSM-IV-TR definition of malingering does not encompass all overreporting, and as will be discussed later, is only one of many competing models of malingering.

Underreporting occurs when an individual's response style leads them to appear better off than they actually are. Individuals may intentionally underreport problems through *minimization* (reporting fewer symptoms than are present and/or minimizing the severity of those that are reported) or *denial* (denying all symptoms) of problems (Ben-Porath, 2003).

Intentional vs. Unintentional Invalid Responding

Most research on overreporting focuses on intentional overreporting (e.g., malingering). However, unintentional overreporting can occur in individuals with poor insight into their problems. More attention has been given to attempts to differentiate between intentional and unintentional <u>under</u>reporting. These have been termed

Table 1

Types of Personality Testing Invalid Responding

Non-Content-Based Invalid Responding

Non-Responding

Random Responding

Intentional Random Responding

Unintentional Random Responding

Fixed Responding

Acquiescence

Counter-Acquiescence

Content-Based Invalid Responding

Overreporting

Intentional Overreporting

Exaggeration

Fabrication

Malingering

Unintentional Overreporting

Low Effort

Underreporting

Intentional Underreporting

Minimization

Denial Impression Management Coaching Practice Effects Self-Deception

impression management and *self-deception*, respectively (Block, 1965; Paulhus, 1984). Impression management occurs when individuals intentionally distort their responses on personality assessment instruments in order to create positive social images, while selfdeception is believed to be an inadvertent or unintentional masking of symptoms (Strong, Greene, Hoppe, Johnston, & Olesen, 1999). Impression management is considered to be a setting-specific strategy, whereas self-deception is thought to be a stable personality trait (Paulhus, 1988).

Misleading Responding Defined

To avoid cumbersome language, the term *misleading responding* is used in this chapter in place of content-based invalid responding. It refers to intentional or unintentional underreporting or overreporting of psychopathology or cognitive complaints. Use of this descriptive term also avoids the use of "malingering" or "defensive," which convey intentionality.

Models of Misleading Responding

A number of models have been proposed that attempt to explain the etiology of some aspect of misleading responding. These models generally fall into three main categories: (a) models that cite underlying psychopathology as a mechanism for misleading responding, (b) models that cite psychopathy, antisocial personality, and criminal behavior as motivation for misleading responding, and (c) models that propose that individuals distort their responses according to a cost-benefit analysis of potential risks and gains associated with their test results. Rogers (1990) labeled these three perspectives as *pathogenic, criminological,* and *adaptational* models of malingering. Additional work has been conducted that expands the models into the realm of underreporting (e.g., Rogers & Dickey, 1991) and unintentional misleading responding (e.g., Tett, Anderson, Ho, Yang, Huang, & Hanvongse, 2006). Research on major models of misleading responding is reviewed and evaluated in this section.

Pathogenic Model of Misleading Responding

Bleuler suggested that overreporting of psychotic symptoms should be seen as a manifestation of mental illness, whether intentional or unintentional (Bash & Alpert, 1980). Rogers (1990) reported that historically, many clinicians have operated under the assumption that malingering is caused by poor coping with an underlying disorder. According to the Pathogenic Model (also called the Coping Model by Heinze, 1999), an individual attempts to assert control over the beginning stages of a chronic mental illness. As the disorder progresses and the individual becomes more impaired, their originally intentionally-produced symptoms develop into unintentional, involuntary symptoms (Rogers & Bender, 2003). Hay (1983) proposed that apparent atypical symptom fabrication could actually be a prodromal stage of psychotic disorder.

The pathogenic model has come under criticism, as the prediction that individuals who are intentionally overreporting symptoms will deteriorate psychologically has not been supported empirically. In fact, the opposite effect has been demonstrated, with many individuals showing marked improvement after their external goal has been reached (Rogers, 1990).

Criminological Model of Misleading Responding

In an effort to become atheoretical, a paradigm shift occurred when the DSM-III (American Psychiatric Association, 1980) introduced a new model of malingering. Labeled alternatively as a "puritanical" and a "criminological" model of malingering (Rogers, 1990; Rogers & Bender, 2003), the DSM model defined malingering as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives," (DSM-III-R, American Psychiatric Association, 1987). According to the DSM-IV-TR (American Psychiatric Association, 2000), any combination of the following should arouse suspicion of malingering:

- Medicolegal context of presentation (e.g., the person is referred by an attorney to the clinician for examination).
- Marked discrepancy between the person's claimed stress or disability and objective findings.

- Lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen.
- 4. The presence of Antisocial Personality Disorder.

Rogers (1990) criticizes this model of malingering, opining that the criteria for diagnosis are generally unrelated except for a common theme of "badness," with "a bad person (sociopath) in a bad situation (forensic assessment), who is a bad participant (lack of cooperation)." Additionally, the DSM-IV-TR does little to address issues in the measurement of intentionality versus unintentionality of overreporting, nor does it address the possibility of malingering in non-forensic settings. Finally, the model does nothing to address intentional underreporting of problems (Rogers & Bender, 2003).

Adaptational Model of Misleading Responding

Rogers and Cavanaugh (1983) introduced the adaptational model of malingering, which assumes that the assessed individual (a) perceives the evaluation process as involuntary or adversarial, (b) perceives they have something to lose from self-disclosure or something to gain from malingering, and (c) does not know of a better way to achieve their goal. It posits that individuals respond in a misleading manner in order to avoid pain, gain financially, or escape punishment (Heinze, 1999). This model provides testable constructs and is presumably free of "moral-laden assumptions," (Rogers, 1990). It seeks to avoid using "mad" (pathogenic) or "bad" (criminological) explanations for malingering and instead frames the construct in terms of predicted utility. Rogers (2008a) notes that most research on response styles has utilized a predictedutility approach to examining the phenomenon. Based on this framework, individuals will be motivated to respond in a misleading manner if they perceive that such a response style will have some utility. For example, an individual undergoing a child custody evaluation would likely find utility in denying their previous substance dependence and exaggerating their moral virtues. Likewise, an individual undergoing a forensic evaluation to determine their sanity at the time of an offense would likely find utility in exaggerating or fabricating symptoms of major mental illness. According to this framework, an individual may overreport some problems while simultaneously underreporting others (Lanyon & Cunningham, 2005; Lees-Haley, English, & Glen, 1991; Rogers, 2008a).

Interpersonal Management Model

The Interpersonal Management Model of deception proposes that individuals exaggerate their psychological symptoms in order to manage their interpersonal relationships. The model is generally focused on misleading responding due to personality pathology. Individuals with Axis II personality traits such as antisocial, dependent, narcissistic, or histrionic characteristics tend to exaggerate their symptomatology in clinical settings. The model rests on an assumption that individuals who overreport symptoms do so because they are "neurotic people whose development was arrested at an early age," (Heinze, 1999). The model does not appear to account for misleading responding in individuals without Axis II psychopathology. It is thus quite limited in scope and is unable to address misleading responding in individuals without personality dysfunction.

Interactional Model of Applicant Faking

Snell, Sydell, and Lueke (1999) introduced an Interactional Model of Applicant Faking to examine the variables that are involved in misleading responding of job applicants. The relatively complex model takes into consideration a variety of explanatory variables in accounting for misleading responding (See Figure 1). It posits that dispositional factors, experiential factors, and test characteristics play a role in determining whether an individual is able to feign symptomatology. At the same time, demographic, dispositional, and perceptual factors will influence the individual's motivation to feign. Both ability and motivation will then play a role in whether individuals will become successful feigners.

The model is very intuitive and rather comprehensive, as it details a number of examples of dispositional factors (e.g., general mental ability, emotional intelligence), experiential factors (e.g., experience on the job, knowledge of job characteristics), and so on, that are hypothesized to play a role in misleading responding on the job.

While much can be gained from the Interactional Model of Applicant Faking, it is proposed to explain only applicant misleading responding, and specifically, underreporting on personality testing and interviewing. A more comprehensive model is needed if we wish to understand both underreporting and overreporting of psychopathology and cognitive complaints across a variety of evaluation settings.



Figure 1. Interactional Model of Applicant Faking (Snell, Sydell, & Lueke, 1999)

Interactionist Model of Item-Level Response Distortion

Tett, Anderson, Ho, Yang, Huang, and Hanvongse (2006) proposed a model of the factors influencing response distortion in job applicants. According to the model, any given test response is determined by a combination of: (a) where the individual falls on the given trait, (b) error due to self-deception, and (c) error due to impression management. Additionally, Tett and colleagues proposed that each of these three components is influenced by three additional factors, including: (a) abilities, knowledge, and skills, (b) targeted and nontargeted personality traits, and (c) situational factors. Figure 2 graphically depicts how each of these variables is hypothesized to interact in order to affect misleading response error in item scores.

Although Tett et al.'s (1995) model was designed to be used within the framework of underreporting within an employment evaluation setting, many of the components are likely to generalize to other evaluation contexts as well as to overreporting of psychopathology. It demonstrates that multiple variables are directly and multivariately involved in the occurrence and successful execution of misleading responding, and separately examines the roles of intentional and unintentional misleading responding on test outcomes. Although the model is promising, more research is needed to determine whether it is empirically supported, and whether it is applicable in nonemployment settings.





Factors Associated with Misleading Responding

Previous research has documented numerous setting and evaluation contexts associated with misleading responding, including inpatient and other medical assessments, drug treatment program evaluations, forensic evaluations, and employment screening. Likewise, numerous individual characteristics have been linked to misleading responding, including demographic variables (i.e., sex, race, age, and educational attainment), cognitive functioning (i.e., intelligence, amnesia, and dementia), and mental disorder (i.e., depression, anxiety, psychosis, eating pathology, addiction, and personality disorders). It is important to understand the associations between these characteristics and the occurrence of misleading responding.

Setting/Context Factors Associated with Misleading Responding

The prototypical individual in a pre-trial forensic assessment for competency to stand trial is likely to exhibit many characteristics that are different than the prototypical individual undergoing an employment screening to become a police officer. Moreover, setting demands influence how individuals emphasize their qualities or problems. For example, an individual who is undergoing a child custody evaluation may be highly motivated to "look their best," both psychologically and cognitively (Bathurst, Gottfried, & Gottfried, 1997), whereas an individual seeking compensation for injuries sustained in an accident stands to gain a great deal of money if they present themselves as both psychologically and cognitively impaired (Berry et al., 1995).

Mental Health Settings. In inpatient medical settings, some individuals are motivated to remain for long periods whereas others wish to be discharged as soon as possible. Regardless of diagnosis, individuals are likely to use impression management to appear in a manner consistent with their goals. Those who wish to remain in treatment are likely to overreport symptoms while those who want to be discharged will underreport (Martin, Hunter, & Moore, 1977).

In Veterans Affairs (VA) medical centers, veterans are treated with medical and mental health services. Additionally, the VA offers service-connected disability compensation for veterans with combat-related Posttraumatic Stress Disorder (PTSD) and other combat-related medical and mental health problems (Frueh, Gold, & Arellano, 1997). Given the potential benefits for individuals with PTSD, there may be a high motivation for some to exaggerate or feign symptoms for service connection and monetary gain. When asked to complete psychological testing, combat veterans reporting PTSD symptomatology tend to produce extreme scores on personality instruments, indicating that they endorse a wide range of severe symptoms (Tolin, Maltby, Weathers, Litz, Knight, & Keane, 2004). Whereas a good portion of these extreme scores can be explained by severity of PTSD symptoms (Hyer et al., 1988), extreme scores produced by some individuals in these evaluations are likely to be due to overreporting of symptoms (Elhai, Gold, Frueh, & Gold, 2000).

Emergency Departments and Substance Treatment Centers. Individual motivations influence self-report accuracy in substance abuse treatment settings and Emergency Department admissions due to substance abuse. Numerous studies have investigated the accuracy of self-reported drug and alcohol use across settings, genders, and races. Most of the literature has found that underreporting of substance use is relatively common, while overreporting is rare in most situations. Underreporting of drug use has been consistently demonstrated in numerous studies, including studies of underreported use of tobacco (Fendrich, Mackesy-Amiti, Johnson, Hubbell, & Wislar, 2005), alcohol (Lapham, C'de Baca, McMillan, & Hunt, 2004; Sommers, Dyehouse, Howe, Wekselman, & Fleming, 2002), marijuana (Fendrich, Mackesy-Amiti, Johnson, Hubbell, & Wislar, 2005; Ghitza, Epstein, & Preston, 2007), opiates (Hser, Maglione, & Boyle, 1999; Morral, McCaffrey, & Iguchi, 2000), and cocaine (Ehrman, Robbins, & Cornish, 1997; Hser, Maglione, & Boyle, 1999; Ledgerwood, Goldberger, Risk, Lewis, & Price, 2008; Lu, Taylor, & Riley, 2001; Messina, Wish, Nemes, & Wraight, 2000).

Overreporting of substance use has been reported in a few studies. Midanik (1982) found that individuals tended to overreport their recent drug use upon intake to a drug and alcohol treatment facility. The author proposed that this unique finding could be attributed to fear of being denied access to the program if one did not appear to be a heavy user. Additionally, overreported use of marijuana has been linked to more frequent use of the drug (Ledgerwood, Goldberger, Risk, Lewis, & Price, 2008).

Forensic Settings. Forensic settings may be some of the most widely-studied in the area of misleading responding. Depending on the type of forensic setting (i.e., criminal court evaluations, correctional evaluations, civil compensation cases, child custody litigation), individuals may have a variety of motivations for distorting responses. Defendants may have very strong motivational forces to distort their response style due to their legal circumstances (Heilbrun, 2001). Research indicates that both underreporting and overreporting may occur in forensic criminal contexts. Some individuals may attempt to appear so disturbed that they cannot be held accountable for the actions with which they are charged. Others may attempt to appear so virtuous that it would be unlikely they would have committed a crime. Still others defendants may

attempt to distort their responses using a combination of both overreporting and underreporting (Dalby, 1988).

Even within the forensic and corrections realm, there are many different motivations for misleading responding that vary across assessed individuals. Walters (2006) outlined seven motivations for exaggeration of psychiatric symptomatology in criminal forensic and correctional settings, including compensation, avoidance, separation, relocation, entitlement, attention, and amusement. Defendants and prisoners may exaggerate problems in order to attempt to sue the prison system for mistreatment, avoid punishment for their actions, be separated from other prisoners, be relocated to a state hospital, attain special privileges reserved for emotionally disordered offenders, gain attention from guards and other inmates, or simply for their own personal amusement.

"Compensation cases," include personal injury, worker's compensation, and disability determination cases (Rogers & Payne, 2006). These cases typically involve an individual claiming the onset of psychological or physical problems, and often involve civil litigation for damages. The assessment process involves determining whether the individual indeed has impaired functioning (Gold, 2004). Individuals may feign in many ways, including overreporting symptoms and features, making false claims that genuine symptoms were caused by the condition in question, and feigning impairment caused by genuine symptoms (Rogers & Payne, 2006). Many studies have demonstrated strong associations between litigation status and overreporting of both psychopathology and cognitive complaints (Berry et al, 1995; Binder & Rohling, 1996; Greiffenstein & Baker,

18

2001; Lanyon & Almer, 2002; Lees-Haley, 1997; Weinborn, Orr, Woods, Conover, & Feix, 2003).

In child custody evaluations, the demand characteristics are very clear: appear competent, virtuous, and high-functioning in order to gain access and/or full custody of one's children. The stakes are very high, and it is not surprising that many parents are motivated to appear well-adjusted. Parents undergoing child custody evaluations tend to score about a standard deviation higher than the normative sample on the MMPI-2 L and K scales, indicating that underreporting of problems and exaggeration of virtues is normative within this setting (Bagby, Nicholson, Buis, Radovanovic, & Fidler, 1999; Bathurst, Gottfried, & Gottfried, 1997).

Employment Settings. Another well-researched setting in misleading responding research is in employment screening. Assessment instruments are often used in employment settings to identify strong job candidates and/or screen out week ones. Most of the research on assessment within this field is in the realm of personality assessment, and in the sub-field, integrity assessment. Integrity assessment involves assessing personality traits as well as attitudes toward theft and past negative behaviors (Jackson, Wrobleski, & Ashton, 2000). Not surprisingly, the competitive nature of the hiring process provides ample motivation for applicants to underreport problems and exaggerate virtues. The extant literature provides solid evidence that underreporting is very common in employment settings. Job applicants have been found to report higher levels of extraversion, conscientiousness, openness, and emotional stability than non-applicants

(Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Griffin, Hesketh, and Grayson, 2004).

Whereas job applicants broadly underreport problems and exaggerate virtues, it appears that the specific personality traits distorted are influenced by job type. In a repeated measures design with a simulation sample, Mahar, Colognon, and Duck (1995) found that people use their knowledge about stereotypical employee characteristics when deciding how to respond on personality instruments in an application context. More research is needed to examine whether applicants would approach personality instruments in different manners when applying for different types of jobs (e.g., electrical engineer versus middle manager positions).

Individual Factors Associated with Misleading Responding

There is much variability in how individuals present themselves within evaluation settings. This is because the unique characteristics of each individual can have a significant influence on their decisions about self-presentation. Characteristics such as sex, race, age, educational attainment, cognitive functioning, and mental disorders have an impact on whether and how individuals may provide misleading information about themselves.

Sex. The association between sex and inaccurate responding is varied, depending on the specific settings investigated. Rosay, Najaka, and Herz (2007) found null differences between the sexes in drug use report accuracy. However, sex differences were found by Klesges, Eck, and Ray (1995), who reported that women were more likely to underreport caloric intake than men in a national health survey of American adults. In most studies of substance abuse reporting, it appears that women are more willing to disclose than are men. Sommers et al. (2002) found that among individuals who suffered injuries in alcohol-related vehicular accidents, men underreported their alcohol consumption more often than women. Differences in underreporting frequencies between the sexes were greater for Caucasians than for African Americans. Women were also less likely than men to underreport their crack cocaine use, according to Lu et al. (2001). In a meta-analysis of socially desirable responding in employment settings, Ones and Viswesveran (1998) found that men tended to score higher on measures of socially desirable responding than women. More systematic investigations of sex and misleading responding across a variety of settings are needed in order to better understand whether and what type of differences exist between the sexes.

Race. Race appears to be an important factor in determining accuracy of selfreport. Much of the literature that has examined race as a predictor of accurate report is in drug and alcohol use research. Many studies have concluded that minority participants were more likely to underreport their true substance use (as verified by urinalysis or hair tests) as compared to Caucasian participants (Fendrich, Mackesy-Amiti, Johnson, Hubbell, & Wislar, 2005; Gray & Wish, 1999; Ledgerwood et al., 2008). Crack cocaine use, however, appears to be underreported in Caucasian and Hispanic individuals more often than in African Americans (Lu et al., 2001). Reports on race differences in marijuana usage differ, with some studies demonstrating that minorities underreport (e.g.,
Ghitza, Epstein, and Preston, 2007), while others demonstrate they overreport (e.g., Ledgerwood, et al., 2008; Rosay et al., 2007) marijuana use, as compared to Caucasian individuals.

Bond and Cherpitel (2004) found that Mexican-Americans in a DUI treatment program were more likely to underreport their BAC levels than their Caucasian counterparts. Aday, Cliu, and Anderson (1980) found that Hispanics tended to be more acquiescent than Caucasian participants, perhaps because of a cultural tendency to want to "please" interviewers from the dominant American culture. Bond and Cherpitel (2004) found that Mexican American participants were more likely to underreport their blood alcohol content levels than Caucasian Americans in a study of individuals involved in alcohol-related motor vehicle accidents. Rates of underreporting were higher for Mexicans born in Mexico than in the United States. Similar results were found by Caetano and Clark (2000).

Research on Asian Americans demonstrates that they are less likely to seek mental health services and more likely to underreport substance use than Caucasians. Researchers believe this is due to cultural beliefs prevalent in many Asian cultures that problems should be kept within the family so as to avoid bringing shame and embarrassment to the entire family (James, Kim, & Moore., 1997; Mercado, 2000). When a family member has substance abuse or other mental health problems, the rest of the family is likely to deny any problem, in order to protect the family's dignity (Ja & Aoki, 1993; James et al., 1997). When race differences in self-report accuracy have been examined, the research has found that minorities tend to underreport drug use more frequently than Caucasians. Hypotheses about these differences have been related to cultural differences regarding shameful behavior or desiring to please Caucasian interviewers. Less research has been conducted to examine racial differences in the veracity of report or effort in other domains, such as personality or cognitive testing.

Age. Most research on misleading responding is conducted on general adult populations. However, some research has examined misleading responding in child and gerontological samples. When research on accurate self-report has been conducted with children, it has often been focused on situations where allegations of sexual abuse or interpersonal violence have occurred. For instance, Bruck, Ceci, and Hembrooke (1998) discuss two problems in child reporting: underreporting of actual abuse, and overreporting of abuse that never happened. Because of the traumatic nature of abuse, avoidance strategies may be developed that lead to underreporting of symptomatology, including dissociation, memory distortion, denial, or emotional numbing. These strategies, while adaptive, can lead to problems in obtaining accurate psychological assessment data. Alternatively, overreporting of problems may be seen as a cry for help (Briere & Elliott, 1997). In extreme situations, children may be likely to fabricate problems if they have been led to believe there is a possibility of secondary gain. Stutts, Hickey, & Kasdan (2003) reported a case of pediatric "malingering by proxy," where a 13-year-old patient feigned somatic problems after being coaxed into malingering by a

parent during a legal settlement. Fortunately, this appears to be a relatively infrequent phenomenon, but more research is needed to determine how often malingering by proxy actually occurs in juveniles and adults alike.

Adolescents may be placed in treatment against their will by parents or court orders. When forced into a treatment in which they are not interested, adolescents are likely to underreport their actual symptomatology in order to distance themselves from therapists or to appear asymptomatic (Baer, Ballenger, & Kroll, 1998). Additionally, they may try to portray themselves in a very favorable light, appearing virtuous as well as asymptomatic (Archer, 2009; Hilts & Moore, 2003).On the contrary, adolescents may also overreport psychopathology in order to avoid criminal responsibility, as a cry for help, or, in an inpatient setting, to avoid being released to a less desirable setting (Stein, Graham, & Williams, 2005). Given that adolescents may underreport or overreport psychopathology, it is important to consider context and motivation in the understanding of adolescent misleading responding.

Research has also examined the link between aging and somatic symptom report, above and beyond the expected increase in somatic and cognitive problems associated with aging. While some believe there is a positive correlation between aging and hypochondriasis (e.g., Butler, 1978), Costa and McCrae (1985) reported data that suggest that increased rates of somatic complaints are simply due to increased problems, and that hypochondriasis is a stable trait that does not lead to more somatic complaints in the elderly. In fact, one concern with the elderly is the *underreporting* of somatic problems.

24

Because degeneration of tissue and nerves leads to decreased pain sensitivity with aging, there is a risk of underreporting of somatic problems due to increased pain thresholds (Basler, 2007).

Additionally, increasing levels of dementia and other cognitive difficulties (discussed in greater detail below) lead to greater difficulty in obtaining accurate selfreport information (Smyth et al., 2002; Basler, 2007). This hypothesis was supported by Rockett, Putnam, Jia, and Smith (2006), who found that elderly participants were more likely to underreport their medication use than younger participants, as verified by toxicology screening. This was attributed to memory difficulties and to being unsure of the chemistry of their medications (i.e., they denied any opiate use, not realizing that their prescription medication was an opiate).

Educational and Vocational Attainment. A few studies have examined the association between educational attainment and invalid responding on testing. Kim and Hill (2003) found that within a sample of young urban African American men, those with higher levels of education were more likely to underreport drug use. The researchers hypothesized that these individuals may have had more to lose from getting in trouble for their drug use and therefore had more motivation to underreport. Similarly, Messina et al. (2000) found an association between employment status and underreporting of post-discharge cocaine use among therapeutic community clients. Specifically, they found that individuals who were employed and not on welfare were more likely to underreport their actual cocaine use, as compared to their counterparts who were unemployed and/or

receiving welfare benefits. It was hypothesized that the employed individuals were afraid to jeopardize their jobs by admitting their use, despite the fact that results were confidential. More research is needed on the association between employment status and underreporting, especially underreporting of psychopathology or cognitive problems.

Intellectual Ability and Disability. Intellectual functioning has been hypothesized to be an important factor in the ability to consciously distort responses on personality instruments. Pauls and Crost (2005) found support for their hypothesis that intelligence is related to both (a) ability to recognize the situational requirements that would influence misleading responding, and (b) ability to recognize the meaning of items on questionnaires, so as to know which items to respond to in an inaccurate manner. In their study, general intelligence was related to both amount of misleading responding and the ability to distort responses in a very specific way so as to "fake a special profile." Mersman and Schulz (1998) found a link between intelligence and mean faking scores and difference scores between honest and faked responses in personality assessment. Intellectual abilities appear to be helpful in understanding both *when* and *how* to distort responses on personality instruments in contexts with different demand characteristics.

Related to social intelligence is the idea of "self-presentation competency," which Hogan, Hogan, and Roberts (1996) describe as the ability to respond in a socially desirable manner. Hogan and colleagues argue that self-presentation competency is related to an awareness of social norms that can be used during the testing process. Pauls and Crost (2005) found support for an association between self-reported self-presentation competency and ability to present oneself in a positive light on personality measures.

Some researchers have examined the topic of Intellectual Disability (ID; formerly referred to as mental retardation) and misleading responding. Pollock (1996) found an association between low intelligence and false-positive reporting on the Structured Interview of Reported Symptoms (SIRS; Rogers, Gillis, Bagby, & Montiero, 1991) and Minnesota Multiphasic Personality Inventory—2nd Edition (MMPI-2; Butcher et al., 2001). Many personality test manuals recommend screening for reading problems before administration, so as to avoid inconsistent and random responding due to poor comprehension (e.g., Butcher et al., 2001). There is a dearth of research that addresses whether individuals with ID have different frequencies or types of misleading responding (i.e., acquiescent responding due to naïveté or gullibility). More research is needed in this area to determine whether there is a link between ID and misleading responding.

Amnesia and Traumatic Brain Injury. Amnesia involves memory deficits in the absence of intellectual and reasoning impairments (O'Connor & Lafleche, 2006). Using simulation designs, van Oorsouw and Merckelback (2004, 2006) examined whether feigning amnesia would have an impact on later recall accuracy. They found that participants who were asked to feign amnesia for a mock event had difficulty accurately recalling true event details when tested one week later. They hypothesized that the best explanation for these findings was likely due to a lack of rehearsal of true information. Similar results were found by Christianson and Bylin (1999) and Bylin and Christianson (2002). Given these simulation design findings, it is possible that individuals who claim amnesia in non-simulation settings may be fabricating memory deficits while *also* experiencing actual difficulty in recall (Kopelman, 2000).

Researchers have also investigated whether having had a traumatic brain injury would affect the ability of individuals to simulate malingering of cognitive impairment. Investigators proposed that individuals who had previous experience with TBI-related impairment would be more successful feigners of cognitive impairment than would controls. In two studies, head-injured patients were no better able to successfully feign neuropsychological deficits than non-clinical controls (Ju & Varney, 2000; Vickery et al., 2004). Ju and Varney (2000) actually found that a history of TBI may have made simulated malingering more difficult because of executive functioning deficits associated with the TBI. Unfortunately, this line of research has not investigated whether having had a TBI would increase the *likelihood* of feigning problems in non-simulation settings. More research is needed to investigate whether there is a link between having had a TBI and invalid symptom report in non-simulation studies.

Dementia. The broad category, "Dementia" includes disorders that are characterized by a deterioration of memory and cognitive abilities. Most research on inaccurate responding in dementia has found that dementia patients underreport dysfunction (e.g., DeBettignies, Mahunn, & Pirozzlo, 1990; Ballard et al., 1991; Green, Goldstein, Sirockman, & Green, 1993). Huntington's disease patients tend to underreport the severity of their difficulties, likely because the disease disrupts brain structures that are involved in awareness (Gontkovsky, 1998). Hoth et al. (2007) found that this lack of self-awareness extended to specific domains, including underreporting problems with behavioral control, emotional control, and activities of daily living. Interestingly, some recent literature has found that overreporting also occurs in individuals with dementia, as compared to caretaker report (Smyth et al., 2002). Basler (2007) noted that pain sensitivity in elderly dementia patients is so poor that self-report is too unreliable and behavioral observations must be used instead.

Depression. Depression symptoms may interfere with accurate self-report of problems. Eaton, Neufeld, Chen, and Cai (2000) reported that underreporting of depressive symptoms was more likely to occur in older respondents, men, and those with less impairment. Similarly, in a sample of elderly depressed inpatients, individuals with less severe physical illness were more likely to overreport somatic symptoms (Duberstein & Heisel, 2007). Depression has also been linked to overreporting of physical functioning disability in older primary care patients (Sinclair, Lyness, King, Cox, & Caine, 2001). Iverson and Binder (2000) discussed the idea that depressed individuals may be misconstrued as intentional feigners because they exhibit cognitive slowing, poor memory, and somatic complaints. However, it may be that these individuals exhibit poor cognitive effort because of their symptomatology, rather than a desire to perform poorly on cognitive testing.

Anxiety. An anxious temperament may have an effect on an individual's perception of negative symptoms and quality of life. Andelman, Fried, and Neufeld

(2001) found that a sample of individuals with left temporal lobe epilepsy tended to be state anxious. This anxious personality was related to overreporting of negative symptoms and poor quality of life, as compared to individuals with right temporal lobe epilepsy who were less anxious.

Trauma and Posttraumatic Stress Disorder. Traumatic experiences can lead to severe psychological symptoms, including avoidance, emotional numbing, cognitive suppression, and increased arousal (Briere & Elliott, 1997). In the most severe cases, this reaction is called Posttraumatic Stress Disorder (PTSD; APA, 2000). Frueh, Hamner, Cahill, Gold, and Hamlin (2000) reviewed the literature on overreporting of PTSD symptomatology. They noted that explanations for symptom overreporting of PTSD in the literature have included a "cry for help," in order to receive treatment and/or attention, malingering to attain disability compensation (Frueh, Gold, & Arellano, 1997), or genuine suffering from the distressing symptoms of the disorder, (e.g., anhedonia or negative affect; Kashdan, Elhai, & Frueh, 2007).

As summarized in a previous section, compensation-seeking status appears to play a significant role in the accuracy of self-reported PTSD (e.g., Frueh et al., 1997; Frueh, Smith, & Barker, 1996; Tolin et al., 2004). Opportunities for significant medical and financial benefits provide incentives for the exaggeration or fabrication of symptoms of PTSD. Veterans at VA Medical Centers may receive financial benefits and/or treatment if diagnosed. Individuals claiming personal injury or filing worker's compensation suits may stand to gain financially if they are believed to have genuine PTSD symptoms. And, criminal defendants may be motivated to feign PTSD as part of a not guilty by reason of insanity defense, to reduce their charges, or to influence their sentencing (Elhai, Gold, Sellers, & Dorfman, 2001). However, when conducting psychological evaluations, clinicians must be careful to not blindly attribute symptoms like inconsistent memory recall, irritability, lack of cooperation, poor concentration, or poor test performance to an intentional desire to feign problems. In fact, these signs that are often used as indices of malingering are genuine symptoms of PTSD (Bordini, Chaknis, Eckman-Turner, & Perna, 2002).

A significant body of literature has investigated the link between PTSD and misleading responding. Most of these studies involve investigating the ability of psychopathology instruments to detect malingered versus genuine PTSD in the context of the Veteran's Administration (VA) setting (e.g., Frueh et al., 1997; Frueh et al., 2000; Kashdan et al., 2007). Because of confounds of the compensation-seeking nature of the VA setting, it is difficult, if not impossible, to separate the influence of PTSD factors versus the compensation-seeking setting on misleading responding. Thus, research on PTSD in settings where secondary gain is absent is more valuable and less confounded with external motivation to feign.

Elhai et al. (2001) compared PTSD outpatients at a childhood sexual abuse survivor treatment center to college students who simulated PTSD on the MMPI-2. Although the outpatients were not screened for litigation status, they had no obvious secondary gain to stand by feigning PTSD symptoms at the treatment center. The PTSD

31

patients scored relatively highly on MMPI-2 indices of overreporting that are confounded with genuine psychopathology, but had only slight elevations on overreporting indices related to very extreme and implausible symptoms. College simulators, on the other hand, scored highly on indices of even extreme, implausible reported symptoms. This suggests that individuals with PTSD and virtually no potential for secondary gain for overreporting problems are not likely to exaggerate their symptomatology to an unrealistic degree.

Other research on trauma and misleading responding has been conducted with sexual abuse survivors. Male survivors of sexual abuse tend to underreport the occurrence of abuse or the resulting problems, in great part due to the stigma of being a male sexual abuse victim (Romano & De Luca, 2001). Geraerts, Jelcic, and Merckelbach (2006) found that individuals who reported recovered memories of childhood sexual abuse did not score higher than other child abuse survivors on the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997) or the Morel Emotional Numbing Test (MENT; Morel, 1998), which are designed to detect misleading responding. Elliott and Briere (1994) found that a sample of children for whom there was substantiated medical or photographic evidence of sexual abuse often denied being abused and, on average, scored lower than non-abused children on the Trauma Symptom Checklist for Children (TSCC; Briere, 1996).

In victimization research, it has been demonstrated that underreporting of physical victimization is a very common phenomenon. The sensitive nature of victimization

32

makes it easier for victims to underreport having been victims rather than to disclose it (Hamby, 2005). Czaja, Blair, Bickart, & Eastman (1994) found that only 29% of people who had previously reported an assault to the police were willing to admit being assaulted to interviewers. Turner (1972) found that 63% of previously-reported victimizations were disclosed to interviewers. More research is needed in samples that are not confounded by motivation to overreport PTSD symptoms (e.g., VA samples, personal injury cases) to better determine whether PTSD symptomatology *itself* may be related to underreporting (e.g., due to avoidance) or overreporting (e.g., as a cry for help) of psychopathology or cognitive problems (Briere & Elliott, 1997).

Schizophrenia and Other Psychotic Disorders. The literature on schizophrenia suggests that patients may be likely to underreport their symptomatology for a number of reasons. Schizophrenia has been linked to long-term memory deficits (Braff et al., 1991; Gruzelier, Seymour, Wilson, Jolley, & Hirsch, 1988), which may have an impact on an individual's ability to accurately report their own symptomatology (Dixon & King, 1995). Additionally, it has been suggested that schizophrenia is linked to anosognosia, which is the minimization or denial of true, debilitating symptoms (Amador, Straus, Yale, & Gorman, 1991). Third, schizophrenic patients may underreport their symptoms out of fear of increased medication or hospitalization (Dixon & King, 1995; Lukoff , Nuecheterlein, & Venura, 1986).

Martin et al. (1977) asked a sample of schizophrenic individuals who had recently been discharged from an inpatient hospital about their overreporting and underreporting. About 36% of the sample admitted to deliberately underreporting symptoms in order to attain a quicker discharge, while an additional 4% admitted to deliberately overreporting symptoms because they were reluctant to leave the hospital. The association between psychotic symptoms and self-report accuracy appears to be complex, motivated by perceived benefits and affected by the disorder itself.

Eating Pathology and Obesity. The literature on eating pathology and misleading responding has been heavily focused on inaccurate reporting of food consumption. Fricker, Baelde, Igoin-Apfelbaum, Huet, and Apfelbaum (1992) found that a subgroup of their sample of obese women underreported their food intake, and that their underreporting was not explained by memory or attention deficits. The authors proposed that this differential report of intake was due to misleading reporting, as the women were reporting what they "should have eaten," rather than what they actually consumed. Additionally, links have been found between consumption misreporting and higher hunger and disinhibition (Lara, Scott, & Lean, 2004). Lara and colleagues also found that consumption underreporting is common in women of all body mass index (BMI) levels, but is more common in women with dieting and binge eating histories and body dissatisfaction. Vansant and Hulens (2006) found an association between restrained eating and underreporting, as well as between emotional eating and overreporting. While this literature may provide clues for hypotheses about the likelihood of individuals with eating pathology to provide misleading responding on measures of psychopathology and

cognitive problems, empirical research with these instruments in such a population is needed to examine whether such links exist.

Pathological Gambling. Individuals who exhibit persistent maladaptive gambling behaviors that interfere with their social and economic functioning may be diagnosed with Pathological Gambling (APA, 2000). Hodgins & Makarchuk (2003) examined the reliability and validity of self-reported gambling behaviors in two samples of problem gamblers. Interestingly, the study found that self-report by gamblers is generally reliable over time and consistent with report by spouses and other collateral contacts. However, these samples may not be representative of all pathological gamblers, as they were all individuals interested in stopping their gambling pursuits. Replications are needed with more representative samples of problem gamblers, including those who are not interested in changing their behaviors. It stands to reason that these individuals are more likely to attempt to conceal their gambling habits from significant others and/or researchers.

Violent Behaviors. Huizinga and Elliott (1986) examined the frequency of underreporting and overreporting of self-reported delinquency in the National Youth Survey (NYS) database. Comparing self-reported delinquency to court records, it was estimated that about 20% of delinquent behaviors went unreported, while an additional 22% to 32% of reported delinquent behaviors were not found in police records (and, therefore, may have been overreported). The estimate of overreporting was less reliable, given that there was the potential for error in documentation of all offenses in the police records. Regardless, these phenomenological data provide significant estimates of

35

misleading reporting of problem behaviors among juvenile delinquents. Archer (1999) conducted a meta-analysis of studies involving the Conflict Tactics Scale (CTS; Straus, 1979), concluding that both men and women tend to underreport their violent behaviors, but that such underreporting appears to be greater among men.

Personality Dysfunction. Research has been conducted examining the association between personality problems, such as neuroticism and antisocial traits, and misleading responding. Costa and McCrae (1985) demonstrated a strong link between neuroticism and number of somatic symptoms reported. Their findings supported the idea that overreporting of symptomatology is related to exaggerations of bodily sensations rather than to actual organic disease. Duberstein and Heisel (2007) found that neuroticism was related to overreporting of emotional and cognitive problems in a sample of elderly depressed patients, as compared to observer report.

Coaching and Practice. Test takers sometimes distort their test responses using coaching. Coaching occurs when an individual is trained by another person (or source, such as the Internet) about how to intentionally distort test responses so as to manipulate validity indices and/or substantive indices on psychological testing (Suhr & Gunstad, 2007). Coaching has been demonstrated to have negative effects on the validity of both personality instruments (e.g., Baer & Sekirnjak, 1997; Lamb, Berry, Wetter, & Baer, 1994; Rogers, Bagby, & Chakraborty, 1993; Rogers, Gillis, et al., 1991; Rogers, Ornduff, & Sewell, 1993; Rogers, Sewell, Morey, & Ustad, 1996) and cognitive instruments (e.g., Coleman, Rapport, Millis, Ricker, & Farchione, 1998; Powell, Gfeller, Hendricks, &

Sharland, 2004; Shum, O'Gorman, & Alpar, 2004; See Suhr & Gunstad, 2007 for a recent review).

Another influence on distorted test results is *practice effects*. Research has demonstrated that individuals generally gain points on intelligence tests with successive attempts due to practice effects (e.g., Basso, Carona, Lowery, & Axelrod, 2002). Thus, those with the opportunity to take multiple intelligence tests are likely to receive higher scores on later administrations because they may remember items or have learned strategies to best find solutions.

Strategies for Detecting Misleading Responding in Personality Assessment

A variety of detection methods have been developed to identify different forms of misleading responding within the realm of personality assessment. This section provides an overview of the approaches and research on their effectiveness.

Detecting Overreported Psychopathology

Overreporting of psychopathology is perhaps the most widely-studied type of misleading responding. Numerous validity scales and stand-alone instruments have been devoted to detecting individuals who exaggerate or fabricate psychiatric symptoms such as psychoticism, depression, or anxiety. A variety of methods have been used in the creation of such validity indices.

Quasi-Rare Symptoms scales are made up of items that are rarely endorsed by normative samples. Although such scales can differentiate between "normal" individuals and those who report severe problems, it can be unclear whether people who endorse

these rare items are doing so because of overreporting or because of genuine psychopathology. *Rare Symptoms* scales address this concern by including items that are rarely endorsed by even clinical samples. These scales are much less confounded by genuine psychopathology than are Quasi-Rare Symptoms scales (Rogers, 2008b). The *Improbable Symptoms* approach is similar to the Rare Items approach, except that items are of an unreal, ridiculous nature. When individuals endorse many of these highly implausible symptoms, it is clear that they are inaccurately reporting their symptomatology. However, these items tend to be so improbable that even moderately sophisticated malingerers can detect and avoid endorsing them (Rogers, 2008b).

The *Symptom Combinations* strategy involves asking about combinations of symptoms where 1) the individual symptoms may commonly occur, but 2) the symptoms rarely occur in combination with each other. Endorsement of one symptom alone is not surprising. Endorsement of the items as a unit, however, leads to a rare symptom combination. The *Spurious Patterns of Psychopathology* method is similar to the Symptom Combinations method, but is based upon scale combinations that have been found with malingerers but not with clinical samples. The *Indiscriminant Symptom Endorsement* approach was created with the assumption that malingerers tend to endorse an overall higher rate of symptoms than do individuals with genuine psychopathology.

Symptom Severity scales operate under the idea that individuals with genuine psychopathology typically do not endorse most or all possible symptoms as present or extreme. These scales measure the severity or number of severe symptoms endorsed.

38

Those who endorse many of these items as severe are likely to be fabricating or exaggerating symptoms. The *Obvious Symptoms* detection method uses face-valid symptoms either alone or in combination with more subtle symptoms to differentiate between malingerers (who are more likely to endorse many symptoms that are obviously related to psychopathology) and those with genuine psychopathology. The *Reported versus Observed* method measures differences between the person's report of their problems and observations made by clinicians, with the assumption that individuals who report multiple problems not seen by clinicians are exaggerating or fabricating symptoms.

The *Erroneous Stereotypes* detection method involves asking about symptoms that individuals often believe are related to true psychopathology, but in fact, are not. Other detection strategies, including *Close Approximations to Genuine Symptoms* and *Overly Specified Symptoms* have been created, but more research is needed to determine whether they are useful strategies. The former relies on the assumption that some feigners report symptoms that are similar to, but not exactly, genuine problems. A concern with this is that individuals with genuine problems may endorse these items and become misclassified as malingerers. The latter method, Overly Specified Symptoms, involves the assumption that malingerers are willing to report problems with greater precision than the average patient. This method lacks much empirical or conceptual support.

Detecting Underreported Psychopathology

A number of detection strategies have been created in order to determine when an individual test taker is minimizing problems or exaggerating positive qualities. Research

on *Response Latencies* has revealed that individuals spend more time answering test items when responding dishonestly (e.g., Holden, 1998; Holden & Kroner, 1992; Holden, Kroner, Fekken, & Popham, 1992). However, this strategy appears to be susceptible to coaching (Robie et al., 2000). Some scales utilize a *Denial of Minor Flaws* method to detect underreporting of psychological problems. Individuals who do not admit to such minor flaws are likely intentionally presenting themselves in a favorable light (Graham, 2006). The *Denial of Psychopathology* method uses items that differentiate between normal controls and individuals with psychopathology who score within normal limits on self-report psychopathology measures. The goal of this method is to distinguish between

Measures for Detecting Misleading Responding

Validity indicators, designed to detect both content-based and non-content-based invalid responding, are found on a number of commonly used self-report personality instruments including the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001) and MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008), Personality Assessment Inventory (PAI; Morey, 1991) and the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, 1994). They also exist as standalone instruments that can be used within larger test batteries, such as the Structured Interview of Reported Symptoms (SIRS; Rogers, 1992), Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005), and Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001).

The MMPI Family

Test developers have been aware of the need for validity indices in personality assessment for decades. Meehl and Hathaway (1946) realized that the original Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1940) should have indices to measure invalid responding. Cannot Say (CNS/?) is a count of the number of items omitted or marked as both true and false by the respondent. If many items are omitted or double-marked, the validity of the scale scores is called into question because these scales will be scored based on incomplete information. CNS provides a simple tool to screen for this form of non-content-based invalid responding as a threat to protocol validity.

The L (Lie) scale was developed to assess defensiveness (underreporting). Developed using the *Denial of Minor Flaws* method, it includes items that describe desirable but uncommon features. The F (Infrequency) scale, developed using the *Quasi-Rare Symptoms* strategy, consists of items that were endorsed by fewer than 10% of the original MMPI normative sample. It was designed to detect respondents who endorsed rare items and was found to be sensitive to non-content-based invalid responding as well as intentional overreporting. The K (Correction) scale was designed using the *Denial of Psychopathology* method and was developed to measure defensive responding that artificially lowered scores on the original MMPI-2 Clinical Scales. K was created by identifying items that differentiated between normal and disturbed individuals who produced within-normal-limits MMPI profiles (McKinley, Hathaway, & Meehl, 1948). Thus, K was able to distinguish between those who were genuinely asymptomatic and those who had symptoms, but denied them.

With the development of the MMPI-2 (Butcher et al., 2001) came additional validity scales. VRIN (Variable Response Inconsistency) is made up of pairs of items with similar or opposite content. Points are scored when individuals endorse these pairs in an empirically and conceptually inconsistent manner. The True Response Inconsistency (TRIN) scale consists of item pairs opposite in content. Points are scored when individuals answer both items in the same direction. It is designed to measure both acquiescence (answering many items in the "True" direction regardless of content) and counter-acquiescence (answering many items in the "False" direction regardless of content). Developed using the *Quasi-Rare Symptoms* approach, the Back Infrequency $(F_{\rm B})$ scale supplements the F scale because it consists of infrequently endorsed items found in the latter part of the MMPI-2 booklet. Infrequency Psychopathology (F_P; Arbisi & Ben-Porath, 1995) was developed using the *Rare Symptoms* approach, as it consists of items infrequently endorsed by both the MMPI-2 normative sample and psychiatric inpatients. Like F, it measures overreporting. However, its design makes it less likely than F to be elevated due to psychopathology, resulting in fewer false positives. The Symptom Validity scale (originally called Fake-Bad Scale; FBS; Lees-Haley, English, & Glenn, 1991) was developed using a rational item selection using the Unusual Symptom *Combinations* approach. FBS was designed to detect malingered emotional distress in

42

individuals undergoing personal injury litigation. It has been shown to be useful in detecting somatic and cognitive overreporting (Ben-Porath, Graham, & Tellegen, 2009).

Seven MMPI-2 validity scales were revised for use with the shorter MMPI-2-RF. VRIN-r consists of item pairs of similar content. Points are earned if the examinee responds in an inconsistent manner (i.e., T/F or F/T). TRIN-r consists of negatively correlated item pairs, and points are earned if the examinee responds to them in the same direction (i.e., T/T or F/F). F-r was created using the *Quasi-Rare Symptoms* strategy; items were selected by identifying ones endorsed by less than 10% of the normative sample.

In order to create maximally distinctive scales, seven F_P items were dropped in the development of F_P -r because they were present on other Validity Scales (i.e., L and F_S). Additionally, correlational analyses led the test developers to remove items with weaker correlations with the full scale score and add a few items with stronger correlations, thus improving the measure. Of the 43 items on the original FBS, 30 were retained in the MMPI-2-RF item pool. FBS-r consists of those remaining 30 items. Infrequent Somatic Responses (F_S) is the sole new validity scale on the MMPI-2-RF, and is designed to identify individuals who are overreporting somatic complaints. The scale, developed using the *Rare Symptoms* strategy, includes items that have somatic content and are relatively uncommonly endorsed by medical patients.

MMPI-2-RF underreporting Validity Scales (including L-r and K-r) were developed after examining a factor analysis of MMPI-2 L, K, S (Superlative SelfPresentation; Butcher & Han, 1995), and Wsd (Wiggins's [1959] Social Desirability) items across various samples. A two-factor solution led to the development of two non-overlapping scales, L-r (with 11 of 14 items coming from L) and K-r (with all 14 items coming from K).

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 1991) contains scales that were designed to detect non-content-based invalid responding, overreporting, and underreporting. The Infrequency (INF) scale, designed to detect careless responding using the *Rare Symptoms* strategy, consists of items rarely endorsed by individuals in normative and clinical samples. Similar to the MMPI-2 VRIN scale, the Inconsistency (INC) scale is designed to detect inconsistent responding and is made up of item pairs of similar or opposite content.

The Negative Impression (NIM) scale was developed using the *Rare Symptoms* strategy, consisting of items rarely endorsed by normative and clinical samples. NIM is designed to identify whether the individual responded in an overly negative manner. However, the scale does not distinguish between intentional overreporting (i.e., malingering) and unintentional exaggeration (i.e., poor insight, crying out for help). Morey (1993) used the *Spurious Patterns of Psychopathology* method to develop the Malingering Index (MAL) which is intended to detect intentional overreporting. Points are scored on the index when an individual profile meets any of eight criteria based on PAI profiles that are more commonly observed in simulated overreporters than honest

responders (Morey, 1996). Rogers and colleagues (1996) developed the Rogers Discriminant Function (RDF) to distinguish between simulators instructed to feign specific disorders and those asked to respond honestly. The RDF consists of 20 PAI indices that best distinguished between the groups (Sellbom & Bagby, 2008).

The original PAI scale developed to detect defensive responding, the Positive Impression (PIM) scale, consists of items endorsed by individuals under instructions to present in a positive manner. Many of the items involve content associated with denial of minor faults. Cashel and colleagues (1995) developed the Cashel Discriminant Function (CDF) to distinguish between honest and defensive responding. The CDF is based on scores on PIM and five other scales (Morey, 1996). Morey (1993) developed the Defensiveness Index (DEF) to detect further assist in the detection of defensive responding. Points are scored on DEF when an individual profile meets any of eight criteria based on PAI profiles that are more commonly observed in simulated underreporters than honest responders (Morey, 1996).

Millon Clinical Multiaxial Inventory - III

The Millon Clinical Multiaxial Inventory (MCMI; Millon, 1983), a personality and psychopathology instrument designed for use in clinical settings, has been revised into the MCMI-II and MCMI-III (Millon, 1987, 1994; Millon, Davis, & Millon, 1997). The MCMI-III includes four validity indices that detect invalid responding and lead to adjustments in substantive scale scores. The first, a Validity Index, consists of three very improbable items. Endorsement of these items suggests that the test-taker was not paying attention while completing the instrument, and the endorsement of two or more items indicates an invalid protocol. The MCMI-III Disclosure scale measures defensive responding (i.e., frankness versus secrecy), whereas the Desirability Index detects overlypositive self-presentations and the Debasement Index was designed to detect exaggeration of problems in functioning (Craig, 2006; Millon, Davis, & Millon, 1997). Millon et al. (1997) described the development of these indices as occurring through a three-step process of (a) rational derivation of items, (b) internal consistency analyses, and (c) external validation analyses. The MCMI-III manual provides only limited information about strategies used to develop these scales.

Structured Interview of Reported Symptoms

The Structured Interview of Reported Symptoms (SIRS; Rogers, Kropp, Bagby, & Dickens, 1992) is a 172-item item structured interview that is designed to detect feigned psychopathology. The scales of the interview were designed using eight different strategies for detection of feigned mental disorders. Scales were developed from rationally-created items, which were refined and developed using expert raters. SIRS Primary Scales include: RS (Rare Symptoms), SC (Symptom Combinations), Improbable or Absurd Symptoms (IA), Blatant Symptoms (BL), Subtle Symptoms (SU), Severity of Symptoms (SEV), Selectivity of Symptoms (SEL), and Reported vs. Observed Symptoms (RO). The test also includes a number of supplementary scales. The SIRS-2 (Rogers, Sewell, & Gillard, 2010) maintained the same set of scales, but made four major modifications, including adding a classification scale (RS-Total, which was designed to differentiate between genuine but atypical and feigned presentations), two indexes, and a supplementary scale.

Structured Inventory of Malingered Symptomatology

The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) is a 75-item self-administered instrument used to detect overreported psychopathology and neuropsychological symptoms in clinical and forensic settings. The instrument includes five scales designed to screen for various forms of malingering. The scales include Psychosis (P; consists of items related to bizarre psychotic symptoms not common in actual psychiatric patients), Neurologic Impairment (NI; consists of items related to highly atypical or illogical neurological problems), Amnestic Disorders (AM; consists of items related to memory impairment not seen in individuals with actual brain injury), Low Intelligence (LI; includes items related to simple, general fund of knowledge), and Affective Disorders (AF; includes items related to a total score that helps determine whether a more complete assessment of malingering is warranted (Smith, 2008).

Miller Forensic Assessment of Symptoms Test

The Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001) is a 25-item structured interview screener that was created to detect overreported psychopathology in forensic settings. The instrument contains seven validity indices created from rationally-constructed items designed to identify a variety of malingering

strategies. For example, the instrument has scales to identify differences between reported and observed symptoms, rare combinations of symptoms, extreme symptoms, and unusual hallucinations. The Reported versus Observed detection strategy was used to develop the RO (Reported vs. Observed) Scale, on which points are scored when individuals report symptoms inconsistent with their behaviors. For example, an item asks the individual whether they have trouble sitting still while giving the examiner a chance to observe such behavior. The ES (Extreme Symptomatology) Scale includes items that measure endorsement of extreme, rare symptoms that are uncommon in general psychiatric patient samples. The RC (Rare Combinations) Scale includes item pairs that are common individually but rare in unison. For example, RC includes an item where the individual is asked (a) whether they have nightmares, and (b) whether those nightmares only occur when they have lost a great deal of weight. The UH (Unusual Hallucinations) Scale utilizes the *Quasi-Rare Symptoms* detection strategy; UH includes items about hallucinations that are so unusual that they are not even endorsed by individuals who have genuine hallucinations. The Unusual Symptom Course (USC) Scale includes one item that addresses whether the individual reports an uncommon symptom course. The Negative Image (NI) Scale includes one item that examines whether the examinee reports an overly-negative view of themselves that is rare in psychiatric populations (with the exception of depressed patients). Finally, the Suggestibility (S) Scale involves a two-part item that assesses whether individuals are susceptible to suggestion. Specifically, at the beginning of the test, the examinee is asked about whether they are experiencing a

symptom that is not actually related to psychopathology, but the examiner suggests that it is. At the end, the examinee is again asked whether they are experiencing that symptom. The M-FAST is designed as a screener for malingering. It is recommended that individuals with elevated scores be more extensively assessed before a determination of malingering is made (Miller, 2001).

Misleading Responding Research Methodology

A variety of research designs have been used to study misleading responding, including simulation, differential prevalence, and known-groups designs. The first of these, *simulation*, involves randomly assigning research participants to follow instructions to respond to test materials in a certain manner. For example, participants may be asked to feign psychopathology or exaggerate virtues. Often, their test scores are compared to clinical, normative, or other samples of individuals who completed the instrument under standard instructions. This experimental design has very strong internal validity because random assignment and experimental conditions can be controlled. However, it has relatively weak external validity because simulators may not respond in a manner similar to actual misleading responders. Thus, findings of these studies may be of limited generalizability to non-simulation populations (Rogers, 2008c).

Differential prevalence research involves comparing test scores of two groups that are presumed to have different levels of misleading responding, typically because one group has a greater incentive to intentionally distort their test responses. For example, scores from a sample of college students may be compared to a sample of forensic pretrial arrestees; the college students are presumed to have no motivation to distort their responses, whereas the arrestees are presumed to have some motivation to exaggerate their psychological problems (i.e., to appear not guilty by reason of insanity). An advantage of this design is that it examines scores from non-simulation samples. However, a significant problem exists with this method: there is no guarantee that all of the individuals in the sample will respond in the expected manner. Thus, some of the individuals in the presumed feigning sample may be responding honestly, and others who may be *motivated* to exaggerate problems may not be successful at doing so. Furthermore, it is possible that some of those individuals could be underreporting psychopathology (i.e., to appear too virtuous to have committed a crime) while others were overreporting problems, thus further muddying any group findings for the sample. In light of these limitations, it is generally not recommended as a strong research design (Rogers, 2008c).

A third commonly-used research design, *known-groups comparison*, involves comparing samples of individuals who can reasonably be labeled as misleading and nonmisleading responders based upon strong external evidence (i.e., documentation that the person is feigning symptoms or failing forced-choice tests at below-chance levels). The major strength of this design is that non-simulation samples can be compared with relatively strong confidence that each sample is made up only of individuals who are either accurate or inaccurate in their responding. This method, however, is not without some limitations. First, because there is no litmus test for invalid responding, there is no

50

complete guarantee that individuals are correctly placed in each group. Thus, this research design is only as strong as the criteria that are used to determine group placement. Secondly, this method assumes that misleading responding is a categorical, rather than continuous phenomenon, as individuals are placed in one (honest) group versus another (feigning) group. In reality, misleading responding is a much more complex phenomenon, with some individuals providing results that are less accurate than others, and others providing a mixture of both underreporting and overreporting. Finally, given that this is an observational design, the approach cannot take advantage of the benefits of experimental design, including random assignment of instructional set conditions (Rogers, 2008c). When well-validated criteria are used for group assignment, the known-groups design is a powerful research tool. When used in conjunction with simulation designs, the two methods are able to complement each other in balancing internal and external validity of findings.

The *Test Validation Summary* technique (Frederick & Bowden, 2008) involves a *mixed-groups validation* method (utilizing mixed samples of honest and misleading responders, where it is unknown which individuals belong to which category) to estimate test classification characteristics (i.e., false positive rate, true positive rate) when presence or absence of feigning is unknown but a *probability* of feigning is estimated. This approach is innovative because it does not have the same sampling requirements as known-groups research (where a high level of confidence that each individual is accurately classified as an honest or misleading responder is required). This relatively

new and promising technique can be used to validate the construct validity of validity indicators (Frederick & Bowden, 2009).

Application of Test Bias Identification Tools in Misleading Responding Research

Test bias occurs when test results have different meanings for different groups. Two main types of test bias exist: internal and external. Internal test bias occurs when two groups score differently on a scale when there is no reason to expect or desire such differences. While internal bias is an important construct, it is outside the scope of this paper and will not be discussed further.

Measuring External Test Bias

External test bias, also known as *predictive bias* or *differential prediction* (Aguinis, Culpepper, & Pierce, 2010) occurs when the association between a test score and a criterion differs between groups or as a function of a continuous moderator variable (Hong & Roznowski, 2001). The most widely-accepted definition of external test bias was developed by Cleary (1968) who explained the concept of test bias as follows:

A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup. With this definition of bias, there may be a connotation of "unfair," particularly if the use of the test produces a prediction that is too low, (p. 115).

In other words, bias can occur if the criterion validity of the test operates in different ways for different groups of individuals. Cleary (1968) goes on to describe how

external test bias may lead not only to differences in predictions of score *values*, but also to differences in the *accuracy* of predictions.

Most commonly, external test bias is examined using moderated multiple regression analyses and can be a result of *slope bias*, *intercept bias*, or a combination of the two (Cleary, 1968). For example, an achievement test (continuous independent variable) may be differentially related to college performance (continuous dependent variable) for males and females (categorical moderator variable). This difference can be in the form of different slopes (indicating differences in the *strength* of the prediction across genders), different intercepts (indicating differences in predicted scores across genders), or both (See Figure 3).

Moderated multiple linear regression can also be used to detect differences across continuous variables (Jaccard & Turrisi, 2003). For example, one may be interested in determining whether an achievement test is equally predictive of college success for individuals at various reading levels (perhaps because poor readers would perform poorly on the exam but could in fact do well in college with student academic services assistance). Rather than using a categorical moderator (i.e., gender), a continuous moderator (i.e., reading score) may be used to examine whether test bias exists across different levels of the moderator variable. Test, slope, and intercept bias are detected using the same moderated multiple regression analyses as previously presented. Plots similar to those presented in Figure 3 may be created by calculating regression lines at various levels of the moderator variable (e.g., at specific low, average, and high reading scores).





While the remainder of the current paper will focus on the use of continuous-level independent and moderator variables, the interested reader is referred to Jaccard and Turrisi (2003) for a more comprehensive review of moderated regression analyses utilizing categorical predictors.

Using a Series of Multiple Linear Regressions. To examine slope and intercept bias when the dependent variable is continuous in nature, a series of multiple linear regression equations are conducted. As described in detail by Aguinis and colleagues (2010), these analyses include the following three regression equations:

- $(1) Y = b_0 + b_1 X + e$
- (2) $Y = b_0 + b_1 X + b_2 G + b_3 X G + e$
- (3) $Y = b_0 + b_1 X + b_2 G + e$

where b_0 is the intercept, b's₁₋₃ are unstandardized coefficients, *G* represents the intercept, *XG* represents the slope, and *e* is random error. Significant differences in R^2 (variance explained) between equations 1 and 2 indicate that some form of test bias is present. Two additional comparisons help to determine which form(s) of test bias is/are present. Significant differences in R^2 between equations 2 and 3 indicate slope bias, whereas differences between equations 1 and 3 indicate intercept bias.

Using Hierarchical Moderated Multiple Linear Regression. Rather than comparing the strength of results across a series of regression equations, it is possible to instead examine *incremental* improvements in model fit across several *blocks* of a single hierarchical moderated regression equation. Because standard statistical packages (e.g., SPSS) produce statistical and practical significance indices (a) at each of the three levels of the equation, and (b) examining the incremental improvement as additional predictors are added to the equation. This method thus readily provides the crucial information needed to examine the strength and direction of moderating effects.

Using Hierarchical Moderated Multiple Logistic Regression. When the dependent variable of interest is binomial in nature, it is most appropriate to use moderated logistic regression models to look for slope and intercept differences. Linear regression is inappropriate with binomial dependent variables (e.g., presence or absence of depression) because this analysis assumes the dependent variable is continuous in nature. Logistic regression, on the other hand, allows for the prediction of the *probability* that the dependent variable is present (e.g., the individual has depression), based upon their scores

on the predictor variables. Statistically, this is achieved by predicting the natural logarithm of the odds of the dependent variable (otherwise called the "log odds" or "logit"), which is denoted as:

(4)
$$L_i = \ln [P_i / (1 - P_i)]$$

where P_i is the *probability* of having the characteristic and $[P_i/(1 - P_i)]$ is the *odds* of having the characteristic. Thus, in logistic regression, there is a predicted *linear* association between the predictors and the *logit* of the dependent variable. In turn, this means there is a curvilinear relationship between the predictors and the predicted dependent variable itself (Pampel, 2000). In the context of moderated logistic regression, an examination of the linear association between the predictors and the logit provide the most direct information about the direction of significant slope and intercept differences (although it is important to be aware that the predicted association with the dependent variable itself is curvilinear).

Using External Test Bias Detection Strategies to Study Misleading Responding Indicators

In a manner similar to those used to detect external test bias, moderated multiple regression analyses can be applied to examine whether validity scales moderate the association between substantive scales and external criteria. In test bias research, slope and intercept bias are problematic; however, in misleading responding research, such differences would demonstrate the ability of validity scales to alert the interpreter to threats to the criterion validity of substantive scale scores. Scores on misleading response

indicators would be expected to moderate the association between substantive scale scores and external criteria.

Applying the framework outlined by Aguinis et al. (2010) to misleading responding indicators would involve the following moderated linear regression equations:

Substantive *Validity + Error

(7) External Criterion = Intercept + Substantive Scale + Validity Scale + Error Of note, the same equations apply in logistic regression, with the exception that the predicted variable is the logit of the External Criterion.

Statistical Significance. In order to examine whether (a) overall moderation, (b) slope differences, and (c) intercept differences are present, it is imperative to examine the *t*-tests that examine incremental improvement across blocks. A statistically significant *t*-test examining the increase in incremental fit from equation 5 to equation 6 indicates that the substantive scale score's criterion validity is moderated by the validity scale score. If a significant result occurs, further analyses are conducted to elucidate the nature of this association.

The next step involves examining whether the *t*-test for the difference between 6 and 7 is significant. A significant finding indicates there are criterion validity slope differences present across levels of the validity scale. In this context, slope differences indicate that the association between the substantive scale and criterion measure differs
across levels of the validity scale, most likely in the direction of lower validity scale scores being associated with higher correlations (although the opposite pattern is possible).

The final step involves examining whether the *t*-test comparing equations 5 and 7 is statistically significant. A significant result indicates that significant criterion validity intercept differences are present across different levels of the validity scale. In this context, intercept differences indicate that the meaning of specific substantive scale scores differ across levels of the validity scale, most likely in the direction of high substantive scale scores being associated with higher levels of the criterion when the validity scale score is lower (although the opposite pattern is possible). A combination of slope and intercept differences is possible, indicating that both (1) criterion validity and (b) the meaning of specific scores differ across validity scale scores.

Practical Significance. Additionally, it is also useful to examine the change in R^2 (variance explained) between equations 5 and 6, as this provides a practically significant indicator of the improvement in model fit. It is also helpful to examine Adjusted R^2 values, as this is a corrected version of R^2 which penalizes the value when additional predictors are used in the model.

Figure 4 demonstrates a hypothetical example of this phenomenon using RC2 (predictor), a psychologist's rating of depression (external criterion), and F-r (moderator). If a moderated multiple regression analysis were applied to these data, scatter plots of the results might look like Figure 4, which illustrates various combinations of slope and

intercept differences at different T score levels on F-r (50 versus 80). Results can be examined to determine whether moderation is present, and whether it is due to slope differences (Panel A; criterion validity is stronger for lower F-r scores than high F-r scores), intercept differences (Panel B; elevations on RC2 are related to higher depression ratings when F-r is low but relatively lower depression ratings when F-r is high), or both (Panel C; a combination of criterion validity and individual score meanings is present).

In this context, slope differences indicate differences in substantive scale criterion validity at varying levels of the validity scale. Whereas this form of moderation is not desired in standard assessment research, it is hoped for in misleading responding research, as it indicates that increasing scores on the validity scale are related to decreases in criterion validity, and thus, that the validity scale is sensitive to invalid responding.

To further elucidate the impact of intercept bias, Figure 5 demonstrates a hypothetical example where the presence of intercept differences leads to differential prediction. In this example, intercept differences mean that a score of 100T on RC2 (Depression) is associated with "high" depression ratings by a psychologist (external, non-self-report criterion) for individuals with normal-range F-r scores (50T), but that 100T on RC2 is actually associated with lower depression ratings for individuals with elevated F-r scores (120T). In other words, the elevation on RC2 is due to actual psychopathology for the honest responders, but is due to exaggeration of true problems for the overreporters. Note also that, in this example, the strength of the association between RC2 and depression ratings can be equal (i.e., r = .40 at both lines). Thus, the



Figure 4. Example of using moderated regression analyses to examine the utility of validity scales to moderate criterion validity of substantive scales.



Figure 5. Example of how intercept differences lead to differential prediction of external criterion scores.

presence of intercept differences alone does not necessarily lead to decrements in criterion validity. Indeed, some research has found that validity scales are robust to very high levels of random responding (Handel, Ben-Porath, Tellegen, & Archer, 2010), which could be explained by the presence of intercept differences. Intercept differences could explain this apparent robustness if individuals with actual psychopathology are uniformly exaggerating their problems rather than fabricating them completely. The correlation between RC2 and depression rating reflects a baseline component of the RC2 score (i.e., pre-exaggeration) that co-varies with the depression rating.

Previous Research on Validity Indices as Moderators of Criterion Validity

A number of studies have examined whether validity indices are able to function as moderators of substantive scale criterion validity. Two types of analyses were conducted in most of these papers: (1) comparisons of substantive scale criterion validity coefficients for groups of valid and invalid responders, or (2) moderated multiple regression analyses, with continuous-level validity scales functioning as moderators of substantive scale criterion validity. Findings from both types of studies will be reviewed.

Non-Content-Based Invalid Responding Validity Scales as Moderators

Hough et al. (1990) created the Assessment of Background and Life Experiences (ABLE) scale, which consisted of four validity scales and ten substantive scales. They administered the ABLE to 9,359 enlisted military personnel who were told their results would be used only for research purposes and would not affect their careers. Although the authors did not elaborate upon the process used, they established a cut score on their

Nonrandom Response Scale (developed to detect non-content-based invalid responding) and compared substantive scale criterion validity for individuals above and below the cut score. They found statistically significant decrements in criterion validity for 19 of 33 (57.58%) correlations between substantive scales and external criteria.

McCrae, Stone, Fagan, and Costa (1998) administered the self-report and spouse ratings forms of the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) to 94 married couples. They examined whether five validity indices, including the Inconsistency (INC) scale (Schinka, Kinder, & Kremer, 1997) moderated the association between self-report and spouse-report rating forms. Although they did not find INC to be a useful moderator, this could have been because the participants were highly educated volunteers who were likely interested enough in the study to answer items in a contentbased manner. Additionally, the authors reported having poor power to detect small effects.

Kurtz and Parrish (2001) also examined the ability of the NEO-PI-R INC scale to moderate NEO-PI-R criterion validity. These authors administered the instrument to undergraduates who enrolled in the study for course credit. Additionally, raters who knew the participant for at least five years (friends, parents, relatives, or romantic partners) completed NEO-PI-R ratings of the participants concurrently and 6 months after the first assessment. Moderated multiple regression analyses demonstrated significant results for only one of seven analyses, with INC moderating the association between NEO-PI-R Extraversion and a measure of gregariousness (increasing variance explained by 2%). However, the direction of the effect was opposite of expected, with higher INC scores indicating stronger criterion validity. The authors acknowledged that their practically insignificant findings could be due to range restriction on the INC variable or the possibility of a non-linear association.

Archer, Fontaine, and McCrae (1998) used three inpatient samples to examine the ability of the MMPI-2 VRIN (Variable Response Inconsistency) scale to moderate substantive scale criterion validity. Specifically, they found that, on average, individuals with VRIN scores \geq 80T tended to produce lower substantive scale correlations with self-and clinician-reported psychopathology ratings. This study provided support for VRIN as a moderator of substantive scale criterion validity, at least using the 80T cut score.

Handel, Ben-Porath, Tellegen, and Archer (2010) used the MMPI-2 non-gendered normative sample and a sample of psychiatric inpatients to examine the impact of random, acquiescent, and counter-acquiescent non-content-based invalid responding on the criterion validity of the Restructured Clinical (RC) scales. Although they did not examine the moderating effects of MMPI-2 or MMPI-2-RF Validity Scales, they were able to examine the moderating effects of increasing percentages of simulated noncontent-based invalid responding. Interestingly, RC Scale criterion validity was not tremendously impacted by random, acquiescent, or counter-acquiescent responding, even up to 70% simulated non-content-based invalid responding. The findings suggest that the RC scales are quite robust to high levels of random, acquiescent, and counter-acquiescent response; however, the findings do not provide information about the utility of the Validity Scales as moderators of substantive scale criterion validity.

Underreporting Validity Scales as Moderators

A number of studies have examined the ability of validity indices to moderate criterion validity in employment settings. Many of these studies produced non-significant findings, suggesting that validity scales may not be able to detect differences in response accuracy in these settings. For example, Hough et al. (1990) found no evidence for the moderating effect of their Social Desirability validity scale on ABLE substantive scale criterion validity in a military sample. Reid-Seiser and Fritzche (2001) found that the NEO-PI-R Positive Presentation Management (PPM; Schinka, Kinder, & Kremer, 1997) scale was not an effective moderator of the association between NEO-PI-R substantive scales and performance ratings in a sample of customer service representatives. Rather, they found PPM to be a useful *substantive* measure of productivity.

However, some investigations have found support for the moderating effects of validity scales in employment settings. For example, Arnold, Feldman, and Purbhoo (1985) found that the Marlowe-Crowne Social Desirability Scale moderated the associations between the predictors, self-reported job satisfaction, commitment, and intention to look for new jobs with the external criterion, job turnover. Holden (2007) used college roommate pairs to examine the ability of the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) Impression Management (IM) scale to moderate substantive scale criterion validity under standard instructions. Although IM was only a

statistically significant moderator for one of five linear regressions, beta weights for all five analyses were in the predicted direction. The author proposed that the lack of motivation to distort responses may have contributed to null findings, as students were paid for participation but had no motivation to answer dishonestly.

A few studies have examined whether underreporting validity scales function as moderators in *non*-employment settings. For example, some researchers used subjects with no motivation to feign in conjunction with partner and friend ratings on the same measures (as external criteria). McCrae, Stone, Fagan, and Costa (1998) did not find moderating effects of the NEO-PI-R Positive Presentation Management (PPM) scale on associations between self-reported and spouse-reported NEO-PI-R substantive scales. Similarly, Piedmont, McCrae, Riemann, and Angleitner (2000) asked volunteers to complete the NEO-PI-R and Multidimensional Personality Questionnaire (MPQ; Tellegen, 1978; 1982) under standard instructions, and volunteers were also rated by peers. MPQ and NEO-PI-R underreporting validity indices did not serve as strong moderators of criterion validity. However, null findings in both of these studies could be attributed to lack of motivation to distort results or to low power to detect small effects.

Also using spouse ratings as external criteria, McCrae and Costa (1983) found that the Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1960) and Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1964) Lie scales functioned better as substantive measures of neuroticism, introversion, and openness than as moderators. Kurtz, Tarquini, and Iobst (2008) administered the Marlowe-Crowne Social Desirability (MCSD) scale and the NEO Five Factor Inventory (NEO-FFI) to college students who received general psychology course credit. The study used roommate and parent NEO-FFI ratings as external criteria, but did not find MCSD to be a significant moderator. In a study examining the association between defensiveness and self-reported anxiety, Eysenck and Derakshan (1999) found that individuals low in social desirability provided more accurate self-ratings of anxiety than did those with higher levels of social desirability.

Only one study has examined the moderating effect of an underreporting validity scale in a correctional setting, where some individuals may have strong external motivations to appear high-functioning. Edens and Ruiz (2006) administered the Personality Assessment Inventory (PAI; Morey, 1991) to 349 inmates, and found a significant moderating effect for the Positive Impression Management (PIM) validity scale in the prediction of institutional misconduct. At present, literature on the utility of underreporting validity scales as moderators of criterion is mixed and limited.

Overreporting Validity Scales as Moderators

Although overreporting occurs in a variety of evaluation contexts and numerous scales have been developed to detect various forms of overreporting, relatively few studies have examined whether existing overreporting validity scales moderate substantive scale criterion validity. McCrae, Stone, Fagan, and Costa (1998) did not find moderating effects of the NEO-PI-R Negative Presentation Management (NPM) scale on associations between self-reported and spouse-reported NEO-PI-R substantive scales. Similarly, Piedmont, McCrae, Riemann, and Angleitner (2000) volunteers completed the NEO-PI-R and Multidimensional Personality Questionnaire (MPQ; Tellegen, 1978; 1982) under standard instructions and were rated by peers. MPQ and NEO-PI-R overreporting validity indices did not serve as strong moderators of criterion validity. However, null findings in both of these studies could be attributed to lack of motivation to distort results or to low power to detect small effects.

Hough, Eaton, Dunnette, Kamp, and McCloy (1990; discussed earlier) found no evidence for the moderating effect of their Poor Impression scale on ABLE substantive scale criterion validity in a military sample. However, participants had no motivation to overreport, and despite being warned that their results were for research purposes only, some participants may have felt that overreporting could lead to problems with promotion.

Despite the widespread popularity of the MMPI family of instruments (Camara et al., 2000), there is no known research that has investigated the ability of MMPI-family overreporting Validity Scales to moderate substantive scale criterion validity. Burchett and Ben-Porath (2010) found that *instructional set* (i.e., overreporting psychopathology or somatic complaints versus standard instructions) has a major effect on MMPI-2-RF substantive scale criterion validity. However, no research has examined whether the *Validity Scales themselves* are able to serve as moderators of criterion validity.

Limitations of Existing Validity Scale Moderation Literature

Based on the reviewed literature, it appears that many studies have found nonsignificant moderating effects of validity scales on substantive scale criterion validity (see Li & Bagger, 2006 and McGrath, Mitchell, Kim, & Hough, 2010 for reviews). However, a number of limitations of these studies could explain why null findings occurred. First, many of the studies were conducted in situations where participants had little or no motivation to distort their responses. In turn, a restriction of range in validity and substantive scale scores could explain a lack of significant moderating effects. Secondly, it should be noted that significant results are predicated upon psychometrically strong 1) validity indices, 2) substantive scales, and 3) criteria, which would serve to generate non-trivial associations. Null findings could be artifacts stemming from weaknesses in any of these areas. Finally, very few studies of moderating effects address whether moderation was due to slope differences, intercept differences, or both. This is a significant limitation, if slope differences are indicative of differences in criterion validity whereas intercept differences are indicative of differential score prediction.

Absent controlled studies utilizing psychometrically strong substantive scales and valid external criteria, null findings cannot be attributed to weaknesses in validity scales. Even further, results can only give information about the utility of the particular validity scale examined, rather than extrapolated to make conclusions about the utility of validity scales in general. Finally, because evaluation setting plays a significant role in motivation to distort responses, it is necessary to examine the utility of validity indices as moderators of criterion validity across a variety of settings.

68

The Current Study

Burchett and Ben-Porath (2010) demonstrated a need for validity indices to detect invalid responding by showing that substantive scale criterion validity is greatly reduced in the presence of over-reporting. Numerous studies have demonstrated that a variety of validity scales are *sensitive to invalid responding*, becoming elevated when individuals are instructed to respond in an invalid manner. However, McGrath, Mitchell, Kim, and Hough (2010) recently questioned the ability of existing validity scales to "enhance the predictive ability of a valid substantive indicator," (p. 452). They proposed that scores on effective validity scales should moderate the validity of substantive scale scores, and concluded that evidence for this is lacking.

This study seeks to examine whether scores on MMPI-2-RF Validity Scales moderate criterion validity of MMPI-2-RF substantive scale scores. The study also seeks to examine whether moderation effects are due to slope differences, intercept differences, or a combination of these phenomena and whether any effects found vary across samples with differing motivations to feign psychopathology. This has major clinical implications, as it would indicate whether MMPI-2-RF Validity Scales indeed give information about when decrements to criterion validity or differences in interpretive meaning are present.

Research Questions

The ability of MMPI-2-RF Validity Scales to moderate criterion validity of MMPI-2-RF substantive scales has never been examined directly. The following research questions were designed to do so.

- Do scores on the MMPI-2-RF overreporting Validity Scales (F-r, F_P-r, F_S, & FBS-r) moderate associations between MMPI-2-RF Restructured Clinical (RC) Scale scores and a variety of conceptually-relevant extra-test indices in simulation and clinical samples where there is motivation to overreport?
 - a. If significant moderation effects are found, are they due (at least in part) to slope differences? In the context of misleading responding research, this would indicate differential substantive scale criterion validity across levels of a continuous validity scale. It was hypothesized that several significant slope differences would be found in the analyses, and that they would indicate that *higher* validity scale scores are associated with *lower* criterion validity. That is, slopes would be flatter at higher validity scale scores (Figure 6, first row).
 - b. If significant moderation effects are present, are they due (at least in part) to intercept differences? In the context of misleading responding research, this would indicate that different levels of *actual* psychopathology are associated with the same *reported* problems across levels of a validity scale. It was hypothesized that several significant intercept differences would be found in the analyses, and that all would be in the direction of *higher* validity scale scores being associated with *lower* intercept values. That is, plotted regression lines of associations between substantive scales and criteria will intersect the intercept at lower values when individuals

obtain higher validity scale scores than when they obtain lower validity scale scores. This would indicate differences in predicted criterion scores, with overreporters having lower predicted actual psychopathology scores (Figure 6, first column).

2. Are MMPI-2-RF overreporting validity scale scores significant moderators of MMPI-2-RF substantive scale criterion validity across a variety of settings where there are varying degrees of motivation to overreport psychopathology? In most extant research on validity scales as moderators of criterion validity, participants have little to no motivation to distort their responses (e.g., studies reviewed by McGrath Mitchell, Kim, & Hough, 2010). This lack of motivation can lead to restriction of range in scores on the self-report indices and less power to detect actual differences. The current study examined moderation effects in four samples where it is assumed that participants had varying degrees of motivation to answer honestly or overreport psychopathology on the MMPI-2-RF. We hypothesized there would be fewer and/or weaker moderator effects in samples with less motivation for misleading responding (i.e., outpatient, inpatient) as compared to samples with more motivation (i.e., college simulators, forensic defendants).



Figure 6. Hypothesized and other graphical depictions of slope and intercept differences.

Method

Participants

Four archival samples were used for the current study. Because demands to distort responses (and ability to attend to the test well enough to provide valid results) can vary tremendously across settings, samples were selected from various types of settings to represent a wide range of psychopathology and reasons for evaluation. All included samples met the following criteria: (1) participants completed the MMPI-2¹ or MMPI-2-RF, (2) non-self-report external criteria were available (with the exception of the college student sample, which included individuals who completed self-report criteria under standard instructions), and (3) there was potential motivation for some individuals to feign psychopathology (i.e., potential for primary or secondary gain, or simulation instructions to do so). In every sample, individuals who responded in a non-content-based invalid manner (CNS (raw) \geq 18, VRIN-r \geq 80T, TRIN-r \geq 80T) were excluded from final analyses.

College Student Controls and Overreporting Simulators (Burchett & Ben-Porath, 2010). Four hundred college students were recruited from introductory psychology courses to complete a series of self-report criterion measures of psychopathological

¹ For the samples where MMPI-2 protocols were administered, item responses were used to score MMPI-2-RF scales. Tellegen and Ben-Porath (2008) demonstrated that MMPI-2-RF protocols scored from MMPI-2 profiles had nearly identical reliabilities and validities as those scored from MMPI-2-RF protocols.

symptoms under standard instructions and then complete the MMPI-2-RF in one of three conditions - (1) under standard instructions, (2) under instructions to feign psychopathology, or (3) under instructions to feign somatic complaints. The third somatic-instructions group was excluded from the current study, leaving a sample of 267 individuals tested under standard instructions (n = 135) or under instructions to overreport psychopathology (n = 132). After excluding individuals who responded to the MMPI-2-RF in a non-content-based invalid manner or who were unable to identify their simulation instructions in a post-test questionnaire (n = 46), the final sample consisted of 221 participants (64 men & 157 women) with a mean age of 19.07 years (SD = 3.05) and a mean of 12.42 years of education (SD = 0.92). Although ethnicity information was not available for this sample, other samples recruited from the same university consisted predominantly of Caucasian participants. Of the four samples, this sample is unique because a simulation study was used, and external criteria consist of *self-reported* responses to a variety of measures that were administered under standard instructions before students were given their experimental instructions (standard or overreporting) to complete the MMPI-2-RF.

Community Mental Health Center Outpatients (Graham, Ben-Porath, & McNulty, 1999). A sample of 1,219 psychiatric outpatients was administered the MMPI-2 during the course of the intake evaluation. Of these outpatients, 826 remained in therapy until at least their fourth session and a Patient Description Form was completed by their treating therapist at that time. After those with non-content-based invalid MMPI-2-RF protocols

were excluded, the final sample consisted of 749 individuals (296 men & 453 women) with a mean age of 33.27 years (SD = 10.28) and a mean of 12.46 years of education (SD = 2.29). The sample included 598 (79.83%) Caucasian individuals, 133 (17.76%) African American individuals, and 18 (2.40%) individuals who self-identified as being from another ethnicity or did not report their race.

Psychiatric Inpatients (Arbisi, Ben-Porath, & McNulty, 2003). A sample of 1,401 psychiatric inpatients from a Veteran's Administration (VA) hospital was combined with a sample of 1,524 psychiatric inpatients from a medical center in a large metropolitan Midwestern city. Patients completed the MMPI-2 during the course of psychiatric treatment. After those with non-content-based invalid MMPI-2-RF protocols were excluded, the final sample consisted of 2,544 individuals (1,916 men & 628 women) with a mean age of 40.37 years (*SD* = 13.80). The combined sample included 2,057 (80.86%) Caucasian individuals, 341 (13.40%) African American individuals, and 146 (5.74%) individuals who self-identified as being from another ethnicity or did not report their race.

Criminal Forensic Assessment Center Defendants (Petroskey, Ben-Porath, & Stafford, 2003). A sample of 2,156 pre-trial criminal defendants was assessed in a Midwestern forensic assessment center that serves municipal and county courts by providing psychodiagnostic evaluations to answer forensic questions. Evaluations include record reviews, psychosocial interviews, and psychological testing, often including the MMPI-2. Of this sample, 1,592 were administered the MMPI-2 as part of their evaluation. The majority of these defendants were evaluated for competency to stand trial (35.30%), sanity at the time of the offense (22.05%), or drug treatment in lieu of conviction (24.31%). After those with missing or non-content-based invalid MMPI-2-RF protocols were excluded, the final sample consisted of 1,358 individuals (996 men & 362 women) with a mean age of 33.51 years (SD = 11.12) and a mean of 11.80 years of education (SD = 2.22). The sample included 1,042 (76.73%) Caucasian individuals, 298 (21.94%) African American individuals, and 18 (1.33%) individuals who self-identified as being from another ethnicity or did not report their race.

Instruments and Measures

Minnesota Multiphasic Personality Inventory—2 *Restructured Form (MMPI-2-RF)*. The MMPI-2-RF (Ben-Porath & Tellegen, 2008) is a 338-item self-report inventory that was developed using a subset of the MMPI-2 item pool. The inventory measures a variety of personality and psychopathology constructs and includes nine Validity Scales² and 42 substantive scales (see Table 2).

² The scope of the current study was limited to examining the moderating effects of the overreporting validity scales (F-r, F_{P} -r, F_{S} , FBS-r) because available archival datasets came from settings where it is expected that at least some participants would be likely to overreport symptoms of psychopathology. At the time of this study, the newest MMPI-2-RF Validity Scale, Response Bias Scale (RBS) was not included on the inventory. Thus, that scale was not included in the current study. Although it would be possible to examine the moderating effects of the validity scales on *all* of the substantive scales of the MMPI-2-RF, the scope of the current study was limited to the Restructured Clinical (RC) scales to avoid producing an unmanageable amount of data. The RC Scales are an ideal set of scales because they measure a variety of domains of psychopathology in a manner that is not extremely specific or extremely broad in scope. RC4

Table 2

The MMPI-2-RF Scales

| Random responding |
|---|
| |
| Fixed responding |
| |
| Responses infrequent in the general population |
| Responses infrequent in psychiatric populations |
| |
| |
| Somatic complaints infrequent in medical |
| patient populations |
| Somatic and cognitive complaints associated at |
| high levels with over-reporting |
| Exaggerated memory complaints |
| Rarely claimed moral attributes or activities |
| Avowals of good psychological adjustment |
| |

(Antisocial Behaviors) was excluded because it is not expected that individuals would exaggerate or fabricate antisocial behaviors.

| | | associated at high levels with under-reporting |
|-----------------------------------|-----|--|
| Higher-Order (H-O) Scales | | |
| Emotional/Internalizing | EID | Problems associated with mood and affect |
| Dysfunction | | |
| Thought Dysfunction | THD | Problems associated with disordered thinking |
| Behavioral/Externalizing | BXD | Problems associated with under-controlled |
| Dysfunction | | behavior |
| Restructured Clinical (RC) Scales | | |
| Demoralization | RCd | General unhappiness and dissatisfaction |
| Somatic Complaints | RC1 | Diffuse physical health complaints |
| Low Positive Emotions | RC2 | Lack of positive emotional responsiveness |
| Cynicism | RC3 | Non-self-referential beliefs expressing distrust |
| | | and a generally low opinion of others |
| Antisocial Behavior | RC4 | Rule breaking and irresponsible behavior |
| Ideas of Persecution | RC6 | Self-referential beliefs that others pose a threat |
| Dysfunctional Negative | RC7 | Maladaptive anxiety, anger, irritability |
| Experiences | | |
| Aberrant Experiences | RC8 | Unusual perceptions or thoughts |
| Hypomanic Activation | RC9 | Over-activation, aggression, impulsivity, and |
| | | grandiosity |
| | | |

Specific Problems (SP) Scales

Somatic Scales

| | Malaise | MLS | Overall sense of physical debilitation, poor |
|----------------------|----------------------------|-----|--|
| | | | health |
| | Gastrointestinal | GIC | Nausea, recurring upset stomach, and poor |
| | Complaints | | appetite |
| | Head Pain Complaints | HPC | Head and neck pain |
| | Neurological Complaints | NUC | Dizziness, weakness, paralysis, loss of balance, |
| | | | etc. |
| | Cognitive Complaints | COG | Memory problems, difficulties concentrating |
| Internalizing Scales | | | |
| | Suicidal/Death Ideation | SUI | Direct reports of suicidal ideation and recent |
| | | | suicide attempts |
| | Helplessness/Hopelessness | HLP | Belief that goals cannot be reached or problems |
| | | | solved |
| | Self-Doubt | SFD | Lack of confidence, feelings of uselessness |
| | Inefficacy | NFC | Belief that one is indecisive and inefficacious |
| | Stress/Worry | STW | Preoccupation with disappointments, difficulty |
| | | | with time pressure |
| | Anxiety | AXY | Pervasive anxiety, frights, frequent nightmares |
| | Anger Proneness | ANP | Becoming easily angered, impatient with others |
| | Behavior-Restricting Fears | BRF | Fears that significantly inhibit normal activities |
| | | | |

| Multiple Specific Fears | MSF | Fears of blood, fire, thunder, etc. |
|---|--------|---|
| Externalizing Scales | | |
| Juvenile Conduct | JCP | Difficulties at school and at home, stealing |
| Problems | | |
| Substance Abuse | SUB | Current and past misuse of alcohol and drugs |
| Aggression | AGG | Physically aggressive, violent behavior |
| Activation | ACT | Heightened excitation and energy level |
| Interpersonal Scales | | |
| Family Problems | FML | Conflictual family relationships |
| Interpersonal Passivity | IPP | Being unassertive and submissive |
| Social Avoidance | SAV | Avoiding or not enjoying social events |
| Shyness | SHY | Bashful, prone to feeling inhibited and anxious |
| Shyness | 5111 | around others |
| Disaffiliativeness | DSF | Disliking people and being around them |
| Interest Scales | | |
| Aesthetic-Literary | | Literature music the theoton |
| Interests | AES | Literature, music, the meater |
| Mechanical-Physical | MEC | Eiving and building things the outdoors sports |
| Interests | WILC | Traing and building unings, the buildbors, sports |
| Personality Psychopathology Five (PSY-5) Scales | | |
| Aggressiveness-Revised | AGGR-r | Instrumental, goal-directed aggression |

| Psychoticism-Revised | PSYC-r | Disconnection from reality |
|------------------------|--------|--------------------------------------|
| Disconstraint-Revised | DISC-r | Under-controlled behavior |
| Negative Emotionality/ | NEGE-r | Anxiety, insecurity, worry, and fear |
| Neuroticism-Revised | | |
| Introversion/Low | INTR-r | Social disengagement and anhedonia |
| Positive Emotionality- | | |
| Revised | | |

Note. Table reproduced from Ben-Porath & Tellegen (2008).

The following instruments and review forms were used as external criteria in the current study. Table 3 provides information about which measures are present in each sample.

Antisocial Behavior Questionnaire (ABQ). The ABQ is a 16-item questionnaire that measures self-reported delinquent behaviors. Developed by Sellbom and Verona (2004), it was modified from a previous self-report delinquency questionnaire (Hirschi, Hindelang, & Weis, 1980; Lynam et al., 1999). The ABQ is strongly correlated with the Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996) and the Impulsive Antisociality PPI subscale (Sellbom & Verona, 2004). Sellbom (2007) reported an ABQ internal consistency level (Chronbach's α) of .81.

Alcohol Use Disorders Identification Test (AUDIT). The AUDIT is a 10-item alcohol abuse screening tool. It has been validated with a variety of samples including drug users, university students, primary care patients, and geriatric hospital patients, and is highly correlated with other measures of alcohol use, including the Michigan Alcohol Screening Test (r = .88) and the CAGE Questionnaire of Alcohol Abuse (r = .78) (Babor et al., 2001).

Cognitive Difficulties Scale – 26-*Item Version (CDS-26).* McNair and Kahn (1983) developed the original 39-item CDS in order to reliably measure memory problems in a self-report format. A 26-item version was created from a subset of CDS items to reduce administration time. The CDS-26 has been validated on 1,628 cognitively healthy adults with normal memory functioning, and is strongly correlated with a 37-item version of the scale (r = .99) (Derouesné et al., 1993).

Drug Abuse Screening Test – 20-Item Version (DAST-20). The DAST-20 is a selfreported drug abuse screening tool (Skinner, 1982), including 20 Yes/No questions. Internal consistency measures (Chronbach's α) have ranged from .74 to .88 in previous research. Additionally, the DAST-20 is highly correlated with other drug abuse measures, including the Addiction Severity Index Drug Composite Score (r = .42) and Alcohol Composite Score (r = .33), as well as the Clinician Rating Scale for Drug Use (r = .40) and the number of days since drug use (r = -.59). DAST-20 sensitivity estimates range from 74% to 89%, while specificity estimates range from 68% to 83% (Yudko et al., 2007).

Emotionality, Activity, Sociability, Impulsivity Temperament Survey – II (EASI-II). The EASI-II (Buss & Plomin, 1975) is a 20-item questionnaire designed to measure four "temperaments." The *Emotionality* scale was designed to assess the intensity of the

Table 3

Instruments and Measures Used in the Present Study

| | Acronym | Sample |
|---|-----------|---------------|
| Test/Measure | | |
| Minnesota Multiphasic Personality | | A 11 ages - 1 |
| Inventory—2 Restructured Form | MMPI-2-RF | All samples |
| Antisocial Behavior Questionnaire | ABQ | College |
| Alcohol Use Disorders Identification Test | AUDIT | College |
| Cognitive Difficulties Scale | CDS | College |
| Drug Abuse Screening Test | DAST | College |
| EASI Emotionality | EASI-E | College |
| EASI Activity | EASI-A | College |
| EASI Sociability | EASI-S | College |
| EASI Impulsivity | EASI-I | College |
| Goldberg Mania Scale | GMS | College |
| General Self-Efficacy Scale | GSE | College |
| Rosenberg Self-Esteem Scale | SES | College |
| SCL-90-R Somatization | SOM | College |
| SCL-90-R Obsessive-Compulsive | O-C | College |
| SCL-90-R Interpersonal Sensitivity | I-S | College |
| SCL-90-R Depression | DEP | College |

| ANX | College |
|------|--|
| HOS | College |
| PHOB | College |
| PAR | College |
| PSY | College |
| GSI | College |
| PSDI | College |
| PST | College |
| FRRF | Forensic Pre-Trial |
| IRRF | Inpatient |
| PDF | Outpatient |
| | ANX HOS PHOB PAR PSY GSI PSDI PSDI FRRF IRRF IRRF PDF |

individual's reactions to stimuli, while the *Activity* scale was designed to assess how energetic the individual reported being and the *Sociability* scale was designed to measure affiliativeness. Finally, the *Impulsivity* scale was designed to measure one's tendency to react quickly to environmental stimuli.

Goldberg Mania Scale (GMS). The GMS is an 18-item questionnaire designed to assess manic symptom severity in patients. Item-total score correlations have been found to range from .35 to .73 (I. Goldberg, personal communication, July 13, 2008). To date, no data have been published on the psychometric properties of the GMS.

General Self-Efficacy Scale (GSE). The GSE is a 10-item questionnaire that was created to measure self-efficacy, which is how *able* individuals perceive themselves to be in broad domains of their lives. Scholz, Doña, Sud, and Schwarzer (2002) found that corrected item-total correlations for the GSE were high in the United States (Chronbach's $\alpha = .87$) and in other countries (Chronbach's α ranged from .75 to .91). Over 1,000 empirical studies have been conducted to validate various versions of the GSE (Schwarzer, 2008; Schwarzer & Born, 1997).

Rosenberg Self-Esteem Scale (SES). The SES is a 10-item self-report measure of how much people believe they are "good enough," (Rosenberg, 1989). It was designed to be a short, unidimensional, face-valid measure of self-esteem.

Symptom Checklist 90 – Revised (SCL-90-R). The SCL-90-R (Derogatis, 1994) is a 90-item self-report instrument that includes nine primary dimensions of mental health problems as well as three global psychopathology indices. The SCL-90-R has been normed on outpatients, inpatients, normals, and adolescent normals. Derogatis (1994) reported that the nine primary dimensions are highly correlated with conceptually-similar MMPI Clinical Scales.

Forensic Record Review Form (FRRF). The FRRF (Petroskey, Ben-Porath, & Stafford, 2003) is a lengthy form that was used to collect collateral data from each defendant's assessment file in the criminal forensic assessment center sample. The FRRF includes information about social, educational, medical and mental health, legal,

employment, and abuse history as well as mental status at the time of the evaluation. Information reviewed to complete the FRRF came from a variety of sources, including prosecutor files documenting alleged offenses, public records regarding previous convictions, treatment records, educational records, a psychosocial history report, and final psychological reports that were written by licensed psychologists to answer forensic questions (i.e., competency to stand trial, sanity at the time of the offense). The FRRF was completed by trained research assistants. Two assistants completed FRRFs for the same 10% of the cases and had mean Kappa values of .77 (for dichotomous variables) and mean intraclass correlation values of .78 (for continuous variables) (Petroskey, Ben-Porath, & Stafford, 2003).

Inpatient Record Review Form (IRRF). The IRRF (Arbisi, Ben-Porath, & McNulty, 2003) was created to obtain standardized patient data from the inpatient samples. The IRRF includes items related to demographic, admission, and previous treatment information as well as diagnoses and legal history. Additionally, the IRRF includes 87 items that addressed mood, affect, and cognitions which were coded by a rater who reviewed patient charts and intake interview records. Finally, the IRRF includes information about discharge summaries, including DSM diagnosis, improvement, disposition, and medications prescribed. However, because MMPI-2 results were available at the time of discharge (and therefore had potential to influence discharge ratings), discharge variables were not used in this study. Patient Description Form (PDF). The PDF (Graham, Ben-Porath, & McNulty, 1999) is a rating form used in the outpatient community mental health sample. The PDF consists of 188 therapist ratings of personality characteristics and symptoms of community mental health center outpatients after the commencement of three therapy sessions. Ratings are on a five-point Likert-type scale, with "1" indicating the description was "Not at all" representative of the client, and "5" indicating that the client was considered "Very high" on the characteristic. Therapists made ratings about their clients while blind to MMPI-2 results, so PDF ratings are independent from scores on the instrument.

Procedures

The university student sample is comprised of students who were solicited from introductory psychology courses and received course credit for participation. Each participant completed a battery of self-report criterion measures under standard instructions and was randomly assigned to complete the MMPI-2-RF either (1) under standard instructions, (2) under instructions to feign psychopathology, or (3) under instructions to feign somatic complaints. The feigned somatic complaints group was excluded from the current study. The feigned psychopathology participants were given the following instructions:

"Instead of following the regular directions for taking the MMPI-2-RF, I'm going to ask you to take the test in a different way. When filling out the MMPI-2-RF, pretend that you have been charged with a serious crime, such as murder. You are facing a life sentence and possibly even the death penalty. However, you are aware that some people have been found not guilty of serious crimes because of

mental health problems. Please take this test as though you were trying to appear as though you had serious mental health problems and that those problems might help you get a lesser sentence or possibly even found not guilty of the murder you committed. But, you want to make sure that it's not so obvious that the test results indicate that you were lying. So, please do your best to look as mentally ill as possible, without being too obvious about it."

Individuals in the community mental health center sample were administered the MMPI-2 during the course of intake evaluations. Additionally, a Patient Description Form was completed by therapists after the third therapy session, rating clients on 188 personality and psychopathology characteristics. Individuals in the combined Veteran's Administration / medical center psychiatric inpatient sample were administered the MMPI-2 during the course of their inpatient treatment. Inpatient record review forms were completed for each patient by trained research assistants after discharge and included recording information about patient demographics, symptoms, diagnoses, mood, affect, cognitions, and legal history based on patient medical chart information. Individuals in the criminal forensic assessment sample were assessed during the course of criminal psychodiagnostic evaluations, such as competency to stand trial and sanity at the time of the offense, among others. Defendants were administered MMPI-2 protocols as part of the evaluation process. Upon completion of evaluations, trained research assistants completed forensic record review forms based upon defendant chart information including prosecutor information, legal records, treatment records, educational records, psychosocial history, and final reports written by psychologists to answer forensic questions.

Operationalization of Study Hypotheses

Research Question One. The proposed study includes two research questions. The first question involves the exploration of MMPI-2-RF overreporting validity scales (F-r, F_P-r, F_S, & FBS-r) as moderators of MMPI-2-RF Restructured Clinical (RC) Scale criterion validity. It was hypothesized that significant moderation effects, including both slope and intercept differences, would be found across RC Scale/Criterion Measure combinations and across samples. Because there is to date no information regarding which of these validity scales might be significant moderators of substantive scale criterion validity, or whether it would occur due to slope effects, interaction effects, or both, our examination of patterns across scale combinations was exploratory in nature. To examine the patterns of significant moderation effects, a series of hierarchical moderated multiple regression analyses were conducted, where the substantive scale, the validity scale, and an interaction between the substantive and validity scale were entered into regression models in a hierarchical fashion. Incremental improvements in model fit indicated whether overall moderation was present and whether moderation was due to slope or intercept differences (as discussed earlier based on Aguinis et al., 2010). Further examination of the regression findings indicated whether (a) the significant slope differences indicated that higher validity scale scores were associated with lower criterion validity, and (b) the intercept differences were in the direction of *higher* validity scale scores being associated with *lower* intercept values, as hypothesized (as depicted in Figure 6).

Operationalization of Research Question One. The selection of scale and criterion measure combinations to use in the regression models was a multi-step process. First, we selected four archival samples with external criterion measures that we identified as being unconfounded by self-report bias. For the college combined simulator and control sample, participants completed self-report criterion measures under standard conditions before approximately half of the sample was asked to complete the MMPI-2-RF under feigned psychopathology instructions. The three non-simulation samples (inpatient, outpatient, forensic) involved individuals completing testing and interviewing in a context that had a significant impact upon their lives. Any biases or response styles an individual has when completing one instrument (e.g., MMPI-2) would likely to be present when completing other questionnaires (e.g., self-report questionnaires about depression, anxiety, etc.). Thus, it would be inappropriate to use self-reported test data as an external measure of a participant's true functioning. We therefore selected samples which had non-self-report external data (chart data, psychologist ratings, etc.) so that we could use those variables as external measures of participants' functioning.

After selecting our samples, we examined the available appropriate external criteria to determine which variables appeared to be conceptually related to any of the RC Scales. We then examined bivariate correlations between RC Scales and conceptually-related criteria (a) on the original sample, (b) after non-content-based invalid profiles were removed, and (c) after non-content and content-based invalid profiles were removed. These correlations were examined in order to determine whether those criteria

that were believed to be conceptually associated to the RC scales were indeed statistically correlated. In order to avoid attributing null findings to poorly-moderating validity scales when there is no association to be moderated, we did not conduct moderated regression analyses on any RC Scale/Criterion Measure pairings that had correlations weaker than |.20| after invalid protocols were removed based on (c) above. The |.20| level was chosen because it is a commonly-utilized level to denote practically significant association in MMPI scale research.

After those RC Scale/Criterion Measure combinations with weakest associations were eliminated, moderated linear and logistic regression analyses were conducted to examine whether the overreporting validity scales moderated associations between the remaining RC Scale/Criterion Measure combinations. We examined statistical findings whether significant moderation effects were due to slope differences, intercept differences, or both. Finally, we examined moderated regression plots to see whether significant slope and intercept differences were in the expected direction, as depicted in Figure 6.

Research Question Two. The second research question involved the exploration of patterns of significant findings across settings with varying degrees of motivation to overreport. It was hypothesized that fewer and/or weaker moderator effects would be found in settings where *fewer* individuals have motivation to overreport (i.e., outpatient setting, inpatient setting) as compared to settings where *many* individuals have motivation to overreport (i.e., college simulation setting, forensic assessment setting).

Operationalization of Research Question Two. A series of chi square analyses were conducted to examine whether frequency of moderation effects differed across samples. Additionally, chi square analyses examined whether certain validity scales were more frequently significant moderators of criterion validity within each sample.

Results

Bivariate Correlations

In order to examine the statistical association between MMPI-2-RF Restructured Clinical (RC) Scales and conceptually relevant criterion measures, bivariate correlations were examined (a) before any invalid protocols were excluded, (b) after non-content-based invalid protocols were excluded, and (c) after non-content and content-based invalid protocols were excluded (Tables 4 - 7). The general trend of these correlations indicated that, as non-content-based and content-based invalid protocols were removed, the correlations strengthened. Thus, the Validity Scales were helpful in identifying and removing protocols with RC Scales that do not provide useful information about genuine psychopathology.

The correlations calculated after non-content and content-based invalid protocols were excluded provide the cleanest information about the true association between the RC Scales and conceptually-relevant criteria, after participants with random, fixed, overreported, and underreported MMPI-2-RF protocols were excluded. Thus, we used these correlations to determine whether there was indeed a statistical association between the RC Scales and criteria. Because it would be imprudent to attempt to moderate a nonexistent association, those RC Scale/Criterion Measure pairs with correlations below |.20| were eliminated from further analyses, as discussed further below.
College Sample. Table 4 includes bivariate correlations in the mixed college standard and simulated psychopathology feigning sample. There was a notable decrease in the number of protocols included across the three sets of correlations (n's = 267, 221, and 131), indicating there were 46 (17.22%) non-content-based invalid protocols and 90 (33.70%) content-based invalid protocols in the original sample. For most of the RC Scale/Criterion Measure correlations, there was a modest increase in the strength of associations when non-content invalid protocols were removed. The increases in correlational strength were larger when content-based invalid protocols were also excluded. For example, the correlation between RCd and SCL-90-R Depression increased from .31 to .33 when non-content-based invalid protocols were removed. It increased again to .73 when content-based invalid protocols were also excluded. Similarly, the correlation between RC1 and SCL-90-R Somatization increased from .15 to .16 and then to .73 as non-content and content-based invalid protocols were excluded, respectively. This pattern is as expected, as approximately half of this sample was asked to complete the MMPI-2-RF while fabricating and exaggerating psychopathology symptoms.

For some RC Scale/Criterion Measure pairs, the strength of correlations did not significantly increase when invalid protocols were excluded. For example, the correlation between RC2 and EASI Emotionality increased from .11 to .13 and then to .17. Similarly, the RC3 and EASI Sociability correlation only modestly strengthened, from -.09 to -.10 and then to -.16, indicating that the conceptually-hypothesized association between the

scales was not statistically significant. Of 18 conceptually-related associations, three did not end up being associated at |.20| after invalid protocols were excluded. That left 15 college sample associations that were strong enough to be tested for moderation using regression analyses.

Outpatient Sample. Table 5 includes bivariate correlations in the outpatient sample. There was a notable decrease in the number of protocols included across the three sets of correlations (n's = 826, 749, and 657), indicating there were 77 (9.32%) non-content-based invalid protocols and 92 (11.13%) content-based invalid protocols in the original sample. The percentage of invalid protocols was lower than in the college sample, indicating that more individuals were willing and able to provide valid results when assessed in an outpatient treatment setting.

Given the relatively lower amount of invalid responding in this sample, it is not surprising that RC Scale/Criterion Measure correlations tended to be fairly strong in the original sample and did not greatly increase as the few invalid responders were excluded. In fact, associations between most of the RC Scales (RCd, RC1, RC2, RC7, & RC9) did not tend to change very much at all. Correlations with RC3 tended to be very small whether or not invalid protocols were excluded, indicating that this dataset did not have criterion measures that were significantly associated with this scale. Interestingly, the correlations for RC6 and RC8 were relatively low to begin with and *decreased* when content-based invalid protocols were excluded. For example, the association between RC8 and a Patient Description Form rating of Hallucinations decreased from .23 to .18 and then to .05 as invalids were excluded. One possible explanation for this phenomenon is that the therapist's ratings of their patient's psychotic symptoms were strongly confounded by the patient's self-reported symptoms. Another possibility is that the Validity Scales are confounded by genuine psychotic psychopathology. Alternatively, this may be explained by the fact that this sample includes very few, if any, individuals with psychotic disorders because those detected at intake were referred to a separate agency before relevant data were collected. Of 49 conceptually-related associations, 27 did not end up being associated at |.20| after invalid protocols were excluded. That left 22 inpatient sample associations that were strong enough that they *could* be tested for moderation using regression analyses.

Inpatient Sample. Table 6 includes point biserial correlations in the inpatient sample. There was a notable decrease in the number of protocols included across the three sets of correlations (n's = 2,925, 2,544, and 2,023), indicating there were 381 (13.03%) non-content-based invalid protocols and 521 (17.81%) content-based invalid protocols in the original sample. The percentage of invalid protocols was lower than in the college sample, indicating that the majority of these individuals who were assessed in an inpatient treatment setting were willing and able to provide valid results. The invalid rate was somewhat higher than in the outpatient sample, indicating that those in the more acute inpatient setting may have had greater difficulty attending to the test or may have been more likely to report extreme levels of symptomatology in hopes of maintaining their inpatient status.

An examination of the point biserial correlations indicates that in the original sample, the majority of correlations were in the .10 to .20 range. Overall, the correlations did not meaningfully increase as invalid responders were excluded. Of 46 conceptually-related associations, 33 did not end up being associated at |.20| after invalid protocols were excluded. That left 13 inpatient sample RC Scale/Criterion Measure pairs that were strong enough to have an association that *could* be tested for moderation using logistic regression analyses.

Forensic Sample. Table 7 includes bivariate and point biserial correlations in the forensic sample. The number of protocols included across the three sets of correlations decreased from 1,592 to 1,358 and then to 1,103 when non-content-based and content-based invalid protocols were removed, respectively. This indicated that there were 234 (14.70%) non-content-based invalid protocols and 255 (16.02%) content-based invalid protocols in the original sample. The percentage of invalid protocols was similar to that in the inpatient setting, indicating that the majority of these individuals who were assessed in a forensic assessment setting were willing and able to provide valid results.

An examination of the bivariate and point biserial correlations indicates that in the original forensic sample, the correlations for RCd, RC1, RC2 were in the .20 to .30 range, whereas correlations for RC7 and RC9 were in the .05 to .10 range. RC6 and RC8 correlations were more variable, ranging from .03 to .20 in the original full sample. Overall, the strength of the correlations did not meaningfully increase as invalid responders were excluded, and in some cases, decreased slightly. Of 23 conceptually-

related associations, 10 did not end up being associated at |.20| after invalid protocols were excluded. That left 13 forensic sample RC-criterion pairs that were strong enough to have an association that *could* be moderated using logistic regression analyses.

College Sample Bivariate Correlations Between MMPI-2-RF Restructured Clinical Scales and Conceptually-Relevant

External Criteria

| MMPI-2-RF | External Criterion | Original Data | Non-Content-Based | Non-Content & |
|-----------|------------------------------|--------------------|--------------------|-------------------|
| RC Scale | | | Invalids Removed | Content-Based- |
| | | | | Invalids Removed |
| | | (<i>N</i> = 267; | (<i>N</i> = 221; | (<i>N</i> = 131) |
| | | Range = 266 – 267) | Range = 220 – 221) | |
| RCd | SCL-90-R Depression | 0.31* | 0.33* | 0.73* |
| RCd | General Self Efficacy Scale | -0.18 | -0.19 | -0.36* |
| RCd | Rosenberg Self Esteem Scale | -0.32* | -0.34* | -0.76* |
| RC1 | Cognitive Difficulties Scale | 0.10 | 0.14 | 0.26 |
| RC1 | SCL-90-R Somatization | 0.15 | 0.16 | 0.59* |
| RC2 | EASI Emotionality | 0.11 | 0.13 | 0.17 |

| MMPI-2-RF | External Criterion | Original Data | Non-Content-Based | Non-Content & |
|-----------|------------------------------------|--------------------|--------------------|-------------------|
| RC Scale | | | Invalids Removed | Content-Based- |
| | | | | Invalids Removed |
| | | (<i>N</i> = 267; | (<i>N</i> = 221; | (<i>N</i> = 131) |
| | | Range = 266 – 267) | Range = 220 – 221) | |
| RC2 | SCL-90-R Depression | 0.13 | 0.16 | 0.51* |
| RC3 | EASI Sociability | -0.09 | -0.10 | -0.16 |
| RC3 | SCL-90-R Hostility | 0.08 | 0.12 | 0.22 |
| RC6 | SCL-90-R Paranoid Ideation | 0.07 | 0.09 | 0.41* |
| RC7 | SCL-90-R Obsessive-Compulsive | 0.15 | 0.18 | 0.40* |
| RC7 | SCL-90-R Interpersonal Sensitivity | 0.21* | 0.24* | 0.54* |
| RC7 | SCL-90-R Anxiety | 0.15 | 0.19 | 0.41* |
| RC7 | SCL-90-R Phobic Anxiety | 0.06 | 0.10 | 0.21 |
| RC8 | SCL-90-R Psychoticism | -0.02 | 0.00 | 0.33* |

| External Criterion | Original Data | Non-Content-Based | Non-Content & |
|----------------------|---|---|---|
| | | Invalids Removed | Content-Based- |
| | | | Invalids Removed |
| | (N = 267; | (<i>N</i> = 221; | (<i>N</i> = 131) |
| | Range = 266 – 267) | Range = 220 – 221) | |
| EASI Activity | -0.03 | -0.05 | 0.14 |
| EASI Impulsivity | 0.17 | 0.18 | 0.32* |
| Goldberg Mania Scale | 0.20* | 0.25* | 0.41* |
| | External Criterion EASI Activity EASI Impulsivity Goldberg Mania Scale | External CriterionOriginal Data(N = 267; Range = 266 - 267)EASI Activity-0.03EASI Impulsivity0.17Goldberg Mania Scale0.20* | External CriterionOriginal DataNon-Content-BasedInvalids RemovedInvalids Removed(N = 267;(N = 221;Range = 266 - 267)Range = 220 - 221)EASI Activity-0.03-0.05EASI Impulsivity0.170.18Goldberg Mania Scale0.20*0.25* |

Note. *p < .001 (Bonferroni corrected $\alpha = .05/18 = .003$).

Outpatient Sample Bivariate Correlations Between MMPI-2-RF Restructured Clinical Scales and Conceptually-Relevant

External Criteria

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|-------------------------------|-------------------|-------------------|------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 826; | (<i>N</i> = 749; | (N = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RCd | Feels overwhelmed | .29* | .31* | .29* |
| RCd | Tearful | .24* | .25* | .28* |
| RCd | Feels gets raw deal from life | .25* | .25* | .23* |
| RCd | Sad | .32* | .33* | .34* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|------------------------------------|-------------------|-------------------|-------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 826; | (<i>N</i> = 749; | (<i>N</i> = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RCd | Feels hopeless | .38* | .39* | .39* |
| RCd | Feels like a failure | .33* | .34* | .32* |
| RCd | Complains of fatigue | .22* | .23* | .22* |
| RCd | Suicidal ideations | .36* | .38* | .37* |
| RC1 | Difficulty concentrating | .30* | .31* | .30* |
| RC1 | Preoccupation with health problems | .42* | .43* | .44* |
| RC1 | Multiple somatic complaints | .40* | .42* | .43* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|---|-------------------|-------------------|-------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | (<i>N</i> = 826; | (<i>N</i> = 749; | (<i>N</i> = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RC1 | Hypochondriacal | .29* | .31* | .32* |
| RC1 | Physical symptoms in response to stress | .36* | .38* | .39* |
| RC2 | Tearful | .22* | .24* | .29* |
| RC2 | Feels gets raw deal from life | .25* | .24* | .22* |
| RC2 | Depressed | .37* | .40* | .40* |
| RC2 | Sad | .30* | .31* | .33* |
| RC2 | Ruminates | .19* | .20* | .20* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|---------------------------|---------------|-------------------|-------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (N = 826; | (<i>N</i> = 749; | (<i>N</i> = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RC2 | Suicidal ideations | .33* | .35* | .35* |
| RC3 | Hostile | .08 | .09 | .08 |
| RC3 | Low frustration tolerance | .09 | .13* | .10 |
| RC3 | Cynical | .02 | .04 | 01 |
| RC3 | Judgmental | .01 | .00 | .00 |
| RC3 | Pessimistic | .14* | .16* | .11 |
| RC3 | Critical of others | .07 | .07 | .07 |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|----------------------------------|---------------|-------------------|------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (N = 826; | (<i>N</i> = 749; | (N = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RC3 | Indirect expression of hostility | 01 | .01 | 01 |
| RC6 | Difficulty trusting others | .10 | .10 | .10 |
| RC6 | Delusional thinking | .19* | .16* | .05 |
| RC6 | Suspicious | .14* | .14* | .09 |
| RC6 | Psychotic symptoms | .18* | .15* | .06 |
| RC6 | Paranoid features | .17* | .15* | .09 |
| RC7 | Anxious | .21* | .24* | .22* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|----------------------|-------------------|-------------------|------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 826; | (<i>N</i> = 749; | (N = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RC7 | Fears losing control | .10 | .12 | .09 |
| RC7 | Worrier | .23* | .24* | .26* |
| RC7 | Irritable | .12* | .12 | .09 |
| RC7 | Many specific fears | .18* | .19* | .18* |
| RC7 | Nervous | .20* | .21* | .20* |
| RC7 | Obsessive | .08 | .07 | .09 |
| RC7 | Angry | .11 | .11 | .09 |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|---------------------|-------------------|-------------------|------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 826; | (<i>N</i> = 749; | (N = 657; |
| | | Range = 778 – | Range = 707 – | Range = 619 – |
| | | 825) | 748) | 656) |
| RC8 | Delusional thinking | .21* | .17* | .11 |
| RC8 | Suspicious | .15* | .15* | .12 |
| RC8 | Psychotic symptoms | .21* | .18* | .10 |
| RC8 | Hallucinations | .23* | .18* | .05 |
| RC9 | Energetic | .07 | .08 | .09 |
| RC9 | Grandiose | .11 | .14* | .13 |
| RC9 | Agitated | .13* | .14* | .08 |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-------------------|--|--|-----------------------------|---|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (N = 826; | (<i>N</i> = 749; | (N = 657; |
| | | Damas 770 | Damag 707 | D ₂ , , , , , , , , , , , , , , , , , , , |
| | | Range = $7/8 -$ | Kange = 707 - 107 | Range = $619 -$ |
| | | Range = 778 – 825) | 748) | 656) |
| RC9 | Accelerated speech | 825) .05 | 748) | 656) .05 |
| RC9 RC9 | Accelerated speech Excitable | 825) .05 .15* | .06 .17* | .05 .15* |
| RC9 RC9 RC9 | Accelerated speech Excitable Impulsive | Range = 778 - 825) .05 .15* .18* | 748) .06 .17* .18* | .05 .15* .15* |

Note. *p < .001 (Bonferroni corrected $\alpha = .05/49 = .001$).

Inpatient Sample Point Biserial Correlations Between MMPI-2-RF Restructured Clinical Scales and Conceptually-Relevant

External Criteria

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|-----------------------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RCd | Admit problem: Suicidal | 0.23* | 0.25* | 0.25* |
| RCd | Admit medication: Antidepressants | 0.19* | 0.21* | 0.20* |
| RCd | Helplessness/Hopelessness | 0.27* | 0.29* | 0.30* |
| RCd | Worthlessness | 0.18* | 0.20* | 0.21* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|-------------------------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RCd | Discouraged | 0.04 | 0.04 | 0.05 |
| RC1 | Stressors: Chronic medical problems | 0.14* | 0.16* | 0.18* |
| RC1 | Stressors: Acute medical problems | 0.06 * | 0.06 | 0.06 |
| RC1 | Chronic pain | 0.21* | 0.22* | 0.20* |
| RC1 | Memory problems | 0.14* | 0.15* | 0.11* |
| RC1 | History of head injury | 0.07* | 0.06 | 0.02 |
| RC2 | Admit problem: Depression | 0.29* | 0.30* | 0.30* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|-----------------------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC2 | Admit problem: Suicidal | 0.19* | 0.20* | 0.19* |
| RC2 | Admit medication: Antidepressants | 0.20* | 0.21* | 0.20* |
| RC2 | Mood: Depressed | 0.29* | 0.30* | 0.26* |
| RC2 | Loss of interest | 0.20* | 0.20* | 0.21* |
| RC2 | Anhedonia | 0.18* | 0.17* | 0.19* |
| RC2 | Suicidal ideation | 0.28* | 0.29* | 0.27* |
| RC2 | Suicidal attempt | 0.19* | 0.18* | 0.16* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|----------------------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC2 | Suicidal plan | 0.20* | 0.20* | 0.17* |
| RC2 | Ruminations | 0.07* | 0.07* | 0.08* |
| RC6 | Admit problem: Psychoses | 0.19* | 0.16* | 0.25* |
| RC6 | Admit medication: Antipsychotics | 0.11* | 0.11* | 0.10* |
| RC6 | Paranoid/Suspicious | 0.20* | 0.18* | 0.23* |
| RC6 | Ideas of reference | 0.17* | 0.18* | 0.19* |
| RC6 | Delusions of reference | 0.07* | 0.08* | 0.09* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|---------------------------------|-----------------------|-----------------|---------------------|
| RC Scale | | Based Invalids Conten | | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC6 | Delusions | 0.16* | 0.14* | 0.21* |
| RC7 | Admit problem: Anxiety | 0.03 | 0.04 | 0.05 |
| RC7 | Admit problem: Increase in PTSD | 0.16* | 0.18* | 0.16* |
| | symptoms | | | |
| RC7 | Admit medication: Anxiolytics | 0.07* | 0.09* | 0.08* |
| RC7 | Mood: Anxious | 0.06 | 0.06 | 0.06 |
| RC7 | Mood: Angry/Irritable | 0.01 | 0.02 | 0.03 |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|--------------------|---------------------|---------------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (<i>N</i> = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC7 | Worry | 0.01 | 0.01 | 0.01 |
| RC7 | Panic | 0.10* | 0.12* | 0.10* |
| RC7 | Compulsions | 0.00 | -0.01 | 0.01 |
| RC7 | Obsessions | -0.01 | -0.01 | 0.00 |
| RC7 | Nightmares | 0.20* | 0.23* | 0.18* |
| RC7 | Flashbacks | 0.19* | 0.22* | 0.18* |
| RC7 | Intrusive thoughts | 0.10* | 0.11* | 0.08* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|----------------------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC7 | Racing thoughts | 0.07* | 0.08* | 0.06 |
| RC8 | Admit problem: psychoses | 0.15* | 0.12* | 0.17* |
| RC8 | Admit medication: Antipsychotics | 0.12* | 0.13* | 0.13* |
| RC8 | Disorganized | 0.04 | 0.03 | 0.08* |
| RC8 | Delusions of reference | 0.04 | 0.04 | 0.04 |
| RC8 | Delusions | 0.08* | 0.06 | 0.07 |
| RC9 | Admit medication: Antimanics | 0.08* | 0.09* | 0.10* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|--------------------|---------------------|-----------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 2,925; | (N = 2,544; | (<i>N</i> = 2,023; |
| | | Range = 2,729 – | Range = 2,369 – | Range = 1,876 – |
| | | 2,925) | 2,544) | 2,023) |
| RC9 | Impulsive | 0.01 | 0.02 | 0.03 |

Note. *p < .001 (Bonferroni corrected $\alpha = .05/46 = .001$).

Forensic Sample Bivariate and Point Biserial Correlations Between MMPI-2-RF Restructured Clinical Scales and

Conceptually-Relevant External Criteria

| MMPI-2-RF | External Criterion | Original Data | Original Data Non-Content- | |
|-----------|--------------------------------------|---------------------|----------------------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 1,592; | (<i>N</i> = 1,358; | (<i>N</i> = 1,103; |
| | | Range = 1,324 – | Range = 1,138 – | Range = 933 – |
| | | 1,592) | 1,358) | 1,103) |
| RCd | Current medication: Antidepressants | 0.21* | 0.24* | 0.22* |
| RCd | Previous medication: Antidepressants | 0.23* | 0.24* | 0.20* |
| RCd | Current suicidal ideation | -0.24* | -0.24* | -0.18* |
| RCd | Mood: Sad/Depressed | 0.23* | 0.25* | 0.23* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|--------------------------------------|---------------------|---------------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 1,592; | (<i>N</i> = 1,358; | (<i>N</i> = 1,103; |
| | | Range = 1,324 – | Range = 1,138 – | Range = 933 – |
| | | 1,592) | 1,358) | 1,103) |
| RC1 | Previous medical problems (Summed)† | 0.20* | 0.21* | 0.29* |
| RC2 | Current medication: Antidepressants | 0.25* | 0.27* | 0.23* |
| RC2 | Previous medication: Antidepressants | 0.26* | 0.27* | 0.22* |
| RC2 | Current suicidal ideation | -0.27* | -0.27* | -0.20* |
| RC2 | Mood: Sad/Depressed | 0.27* | 0.28* | 0.27* |
| RC6 | Current medication: Antipsychotics | 0.21* | 0.21* | 0.17* |
| RC6 | Previous medication: Antipsychotics | 0.20* | 0.20* | 0.17* |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & | |
|-----------|-------------------------------------|---------------------|---------------------|---------------------|--|
| RC Scale | | | Based Invalids | Content-Based- | |
| | | | Removed | Invalids Removed | |
| | | | | | |
| | | (<i>N</i> = 1,592; | (<i>N</i> = 1,358; | (<i>N</i> = 1,103; | |
| | | Range = 1,324 – | Range = 1,138 – | Range = 933 – | |
| | | 1,592) | 1,358) | 1,103) | |
| RC6 | Thought processes: Delusional | 0.13* | 0.14* | 0.21* | |
| RC6 | Hallucinations | 0.12* | 0.11* | 0.04 | |
| RC6 | Persecutory ideation | 0.11* | 0.10* | 0.09 | |
| RC7 | Current medication: Anxiolytics | 0.05 | 0.06 | 0.06 | |
| RC7 | Previous medication: Anxiolytics | 0.10* | 0.11* | 0.10* | |
| RC8 | Current medication: Antipsychotics | 0.21* | 0.21* | 0.19* | |
| RC8 | Previous medication: Antipsychotics | 0.20* | 0.19* | 0.14* | |

| MMPI-2-RF | External Criterion | Original Data | Non-Content- | Non-Content & |
|-----------|--------------------------------|---------------------|---------------------|---------------------|
| RC Scale | | | Based Invalids | Content-Based- |
| | | | Removed | Invalids Removed |
| | | | | |
| | | (<i>N</i> = 1,592; | (<i>N</i> = 1,358; | (<i>N</i> = 1,103; |
| | | Range = 1,324 – | Range = 1,138 – | Range = 933 – |
| | | 1,592) | 1,358) | 1,103) |
| RC8 | Thought processes - Delusional | 0.03 | 0.03 | 0.01 |
| RC8 | Hallucinations | 0.11* | 0.10* | 0.05 |
| RC8 | Persecutory ideation | 0.04 | 0.04 | -0.02 |
| RC9 | Current medication: Lithium | 0.06 | 0.06 | 0.06 |
| RC9 | Previous medication: Lithium | 0.07 | 0.10* | 0.09 |
| | | | | |

Note. *p < .001 (Bonferroni corrected $\alpha = .05/23 = .002$). † The Previous medical problems (summed) variable is continuous, with associated bivariate correlations. All other correlations in this table are point biserial correlations, as the external criteria are binomial.

Moderated Linear and Logistic Regression Analyses

Tables 8 through 11 report results from a series of linear and logistic regression analyses completed for the four samples. These regressions were conducted to examine whether MMPI-2-RF overreporting Validity Scales (F-r, F_P-r, F_S, FBS-r) moderated RC Scale criterion validity, utilizing RC Scale/Criterion Measure pairs which were identified as meaningfully related in the previously discussed correlational analyses. Although it is often recommended in the literature to utilize standardized or centered scale scores for ease of interpretability (e.g., Jaccard & Turrisi, 2003), it was determined that using raw scores would allow the most interpretable findings in the current analyses. This is because, in MMPI scale research, raw scale scores have greater interpretive meaning than do centered or standardized scores. Further, the use of raw scores allows an easier examination of intercept differences, as the intercept occurs at a raw RC Scale score of zero. For all regressions, Bonferroni-corrected alpha levels³ were utilized to control family-wise error.

College Sample. Because all of the criterion measures in the mixed college standard and simulated psychopathology feigning sample were continuous in nature, all of the moderated regressions for this sample were examined using multiple linear regression analyses (Table 8). Of the 60 moderated regressions examined, 38 (63.33%)

³ The Bonferroni correction was applied separately for each Validity Scale. For example, in the college sample, there were a total of 60 regression analyses, with each of the four Validity Scales moderating 15 associations. Thus, the Bonferroni correction was applied by calculating .05/15 = .003.

were statistically significant at a conservative Bonferroni-corrected alpha of .05/15 = .003. F-r and F_P-r were each significant moderators in 11 (73.33%) of 15 regressions, whereas F_s significantly moderated nine (60.00%) of 15 equations and FBS-r significantly moderated seven (46.67%) of 15 equations. An examination of the 38 significant overall moderation equations indicated that 29 (76.31%) included significant slope effects. All 29 (100.00%) of the significant slope effects were in the expected direction, with lower Validity Scale scores leading to stronger slopes than higher Validity Scale scores. A total of 19 (50.00%) of the 38 significant moderations had significant intercept effects. Of those, only 5 (26.31%) were in the expected direction (as presented in Figure 6), with lower Validity Scale scores leading to higher intercept scores.

Across all of the college sample regressions, the average change in variance explained (ΔR^2) was .0860, indicating that an average of 8.60% additional variability in the predicted criterion measure values was accounted for by adding a Validity Scale moderator (regardless of whether the association was in the predicted direction). When only statistically significant moderations were examined, the average change in variance explained increased to 12.11%.

Within the college sample, the average change in variance explained (ΔR^2) was examined for each Validity Scale separately. For F-r, the average change in variance explained was 10.07%, which increased to 13.00% when only statistically significant moderations were examined. When only significant moderations were examined, the average change in variance explained for F_P-r increased from 9.60% to 12.27%, whereas

123

the average change in variance explained for F_S increased from 7.53% to 10.78% and for FBS-r increased from 7.20% to 12.14%.

Outpatient Sample. All of the criterion measures in the outpatient sample were continuous in nature, so all outpatient moderated regressions were examined using multiple linear regression analyses (Table 9). Of the 88 moderated regressions examined, 36 (40.90%) were statistically significant at a conservative Bonferroni-corrected alpha of .002. F_P-r and F_S each significantly moderated seven (31.81%) of 22 regressions, whereas F-r and FBS-r each significantly moderated 11 (50.00%) of 22 equations. An examination of the 36 significant overall moderation equations indicated that 18 (50.00%) included significant slope effects, all of which were in the expected direction. A total of 21 (58.33%) of the 36 significant moderations had significant intercept effects, but only one of those (4.76%) was in the expected direction.

Across all of the outpatient sample moderated regressions, the average change in variance explained (ΔR^2) was .0154, indicating that an average of 1.54% additional variability in the predicted criterion measure values was accounted for by adding a Validity Scale moderator (regardless of whether the association was in the predicted direction). When only statistically significant moderations were examined, the average change in variance explained increased to 2.81%.

Within the outpatient sample, the average change in variance explained (ΔR^2) was examined for each Validity Scale separately. Results were more modest than for the college sample. For F-r, the average change in variance explained was 1.86%, which

124

increased to 2.64% when only statistically significant moderations were examined. When only significant moderations were examined, the average change in variance explained for F_{P} -r increased from 1.23% to 2.43%, whereas the average change in variance explained for F_{S} increased from 1.00% to 2.29% and for FBS-r increased from 2.09% to 3.36%.

Inpatient Sample. Because all of the criterion measures in the inpatient sample were binomial in nature, all of the moderated regressions for this sample were examined using multiple logistic regression analyses (Table 10). Of the 52 moderated regressions examined, 31 (59.62%) were statistically significant at a conservative Bonferroni-corrected alpha of .004. F-r was a significant moderator in six (46.15%) of 13 regressions, while F_P-r significantly moderated seven (53.85%) regressions. F_S and FBS-r each significantly moderated nine (69.23%) of 13 logistic regression equations. An examination of the 31 significant overall moderation equations indicated that 14 (45.16%) included significant slope effects. A total of 10 (71.43%) of the significant slope effects were in the expected direction, with lower Validity Scale scores leading to stronger slopes than higher Validity Scale scores. A total of 29 (93.55%) of the 31 significant moderations had significant intercept effects. Of those, 22 (75.86%) were in the expected direction, with lower Validity to higher intercept scores.

Across all of the inpatient sample moderated regressions, the average change in variance explained (Δ Nagelkerke R²) was .0254, indicating that an average of 2.54%

additional variability in the predicted log odds of the criterion measure values was accounted for by adding a Validity Scale moderator (regardless of whether the association was in the predicted direction). When only statistically significant moderations were examined, the average change in variance explained increased to 4.10%.

Within the inpatient sample, the average change in variance explained $(\Delta NagelkerkeR^2)$ was examined for each Validity Scale separately. For F-r, the average change in variance explained was 2.92%, which increased to 6.17% when only statistically significant moderations were examined. When only significant moderations were examined, the average change in variance explained for F_P-r increased from 1.85% to 3.14%, whereas the average change in variance explained for F_S increased from 2.15% to 3.00% and for FBS-r increased from 3.23% to 4.56%.

Forensic Sample. One forensic sample criterion (Previous Medical Problems Summed) was continuous, so a set of four moderated linear regressions was conducted to examine the moderating effect of the four Validity Scales on its association with RC1. The remainder of the forensic sample criteria were binomial, so several sets of logistic regressions were examined (Table 11). Of the 36 moderated linear and logistic regressions examined, 14 (38.89%) were statistically significant at a conservative Bonferroni-corrected alpha of .006. F-r was a significant moderator in three (33.33%) of nine regressions, while F_P -r and F_S were each significant moderators in 2 (22.22%) of 9 regressions. Meanwhile, FBS-r significantly moderated seven (77.78%) of nine equations. An examination of the 14 significant overall moderation equations indicated that four (28.57%) included significant slope effects, all of which were in the expected direction. A total of 13 (92.86%) of the 14 significant moderations had significant intercept effects. Of those, eight (61.54%) were in the expected direction.

Across all of the forensic sample moderated regressions, the average change in variance explained ($\Delta R^2 \& \Delta N$ agelkerke R^2) was .0192, indicating that an average of 1.92% additional variability in the predicted criterion measure values (and, in the case of the logistic regressions, the predicted *log odds of* the criterion measure values) was accounted for by adding a Validity Scale moderator (regardless of whether the association was in the predicted direction). When only statistically significant moderations were examined, the average change in variance explained increased to 4.29%.

Within the forensic sample, the average change in variance explained ($\Delta R^2 \& \Delta N$ agelkerke R^2) was examined for each Validity Scale separately. For F-r, the average change in variance explained was 2.22%, which increased to 6.33% when only statistically significant moderations were examined. When only significant moderations were examined, the average change in variance explained for F_P-r increased from 1.89% to 5.50% , whereas the average change in variance explained for F_S increased from 1.33% to 5.50% and for FBS-r increased from 2.40% to 3.00%.

Significance Levels for College Sample Moderated Linear Regression Equations, Slope Differences, and Intercept Differences, and Associated R² Values (N's Range From 220 to 221)

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|-----------------------------|------------|--------------|--------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | (<i>p</i>) | (<i>p</i>) | Model ¹ | Model ² | |
| F-r | RCd | SCL-90-R Depression | <0.001* | <0.001*† | < 0.001* | 0.10 | 0.31 | 0.22 |
| F _P -r | RCd | SCL-90-R Depression | <0.001* | <0.001*† | < 0.001* | 0.10 | 0.30 | 0.20 |
| F_S | RCd | SCL-90-R Depression | <0.001* | <0.001*† | <0.001* | 0.10 | 0.26 | 0.16 |
| FBS-r | RCd | SCL-90-R Depression | <0.001* | <0.001*† | 0.030 | 0.10 | 0.31 | 0.21 |
| F-r | RCd | General Self Efficacy Scale | 0.001* | 0.246 | <0.001*† | 0.03 | 0.09 | 0.06 |
| F _P -r | RCd | General Self Efficacy Scale | 0.001* | 0.153 | < 0.001* | 0.03 | 0.09 | 0.06 |
| F_S | RCd | General Self Efficacy Scale | 0.001* | 0.438 | <0.001*† | 0.03 | 0.09 | 0.06 |
| FBS-r | RCd | General Self Efficacy Scale | 0.001* | 0.003*† | 0.415 | 0.03 | 0.09 | 0.06 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted R^2 Value | | ΔR^2 |
|-------------------|-------|------------------------------|--------------|-------------|-------------|----------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| F-r | RCd | Rosenberg Self Esteem Scale | < 0.001* | <0.001*† | <0.001* | 0.11 | 0.32 | 0.22 |
| F _P -r | RCd | Rosenberg Self Esteem Scale | <0.001* | <0.001*† | <0.001* | 0.11 | 0.34 | 0.24 |
| Fs | RCd | Rosenberg Self Esteem Scale | < 0.001* | 0.002*† | <0.001* | 0.11 | 0.26 | 0.15 |
| FBS-r | RCd | Rosenberg Self Esteem Scale | <0.001* | <0.001*† | 0.025 | 0.11 | 0.32 | 0.22 |
| F-r | RC1 | Cognitive Difficulties Scale | 0.082 | | | 0.02 | 0.03 | 0.02 |
| F _P -r | RC1 | Cognitive Difficulties Scale | 0.045 | | | 0.02 | 0.04 | 0.03 |
| F_S | RC1 | Cognitive Difficulties Scale | 0.030 | | | 0.02 | 0.04 | 0.03 |
| FBS-r | RC1 | Cognitive Difficulties Scale | 0.046 | | | 0.02 | 0.04 | 0.03 |
| F-r | RC1 | SCL-90-R Somatization | < 0.001* | <0.001*† | <0.001*† | 0.02 | 0.22 | 0.21 |
| F _P -r | RC1 | SCL-90-R Somatization | <0.001* | <0.001*† | <0.001*† | 0.02 | 0.26 | 0.25 |
| Fs | RC1 | SCL-90-R Somatization | < 0.001* | <0.001*† | <0.001*† | 0.02 | 0.19 | 0.18 |
| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|----------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| FBS-r | RC1 | SCL-90-R Somatization | < 0.001* | <0.001*† | 0.641 | 0.02 | 0.11 | 0.10 |
| F-r | RC2 | SCL-90-R Depression | < 0.001* | <0.001*† | 0.080 | 0.02 | 0.14 | 0.12 |
| F _P -r | RC2 | SCL-90-R Depression | < 0.001* | <0.001*† | 0.014 | 0.02 | 0.13 | 0.11 |
| Fs | RC2 | SCL-90-R Depression | < 0.001* | <0.001*† | 0.192 | 0.02 | 0.11 | 0.10 |
| FBS-r | RC2 | SCL-90-R Depression | < 0.001* | <0.001*† | 0.074 | 0.02 | 0.11 | 0.10 |
| F-r | RC3 | SCL-90-R Hostility | 0.056 | | | 0.01 | 0.03 | 0.03 |
| F _P -r | RC3 | SCL-90-R Hostility | 0.159 | | | 0.01 | 0.02 | 0.02 |
| F_S | RC3 | SCL-90-R Hostility | 0.058 | | | 0.01 | 0.03 | 0.03 |
| FBS-r | RC3 | SCL-90-R Hostility | 0.117 | | | 0.01 | 0.02 | 0.02 |
| F-r | RC6 | SCL-90-R Paranoid Ideation | < 0.001* | <0.001*† | 0.067 | 0.00 | 0.11 | 0.11 |
| F _P -r | RC6 | SCL-90-R Paranoid Ideation | 0.001* | <0.001*† | 0.760 | 0.00 | 0.06 | 0.07 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|------------------------------------|--------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| Fs | RC6 | SCL-90-R Paranoid Ideation | 0.001* | 0.001*† | 0.235 | 0.00 | 0.05 | 0.06 |
| FBS-r | RC6 | SCL-90-R Paranoid Ideation | 0.012 | | | 0.00 | 0.04 | 0.04 |
| F-r | RC7 | SCL-90-R Obsessive-Compulsive | < 0.001* | 0.001*† | 0.019 | 0.03 | 0.09 | 0.07 |
| F _P -r | RC7 | SCL-90-R Obsessive-Compulsive | 0.001* | 0.005 | 0.010 | 0.03 | 0.08 | 0.06 |
| F_S | RC7 | SCL-90-R Obsessive-Compulsive | 0.003* | 0.004 | 0.061 | 0.03 | 0.07 | 0.05 |
| FBS-r | RC7 | SCL-90-R Obsessive-Compulsive | 0.007 | | | 0.03 | 0.06 | 0.04 |
| F-r | RC7 | SCL-90-R Interpersonal Sensitivity | <0.001* | <0.001*† | <0.001* | 0.05 | 0.17 | 0.12 |
| F _P -r | RC7 | SCL-90-R Interpersonal Sensitivity | <0.001* | <0.001*† | <0.001* | 0.05 | 0.17 | 0.12 |
| F_S | RC7 | SCL-90-R Interpersonal Sensitivity | < 0.001* | <0.001*† | 0.001* | 0.05 | 0.15 | 0.11 |
| FBS-r | RC7 | SCL-90-R Interpersonal Sensitivity | < 0.001* | <0.001*† | 0.523 | 0.05 | 0.13 | 0.09 |
| F-r | RC7 | SCL-90-R Anxiety | 0.001* | 0.024 | 0.003* | 0.03 | 0.09 | 0.06 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|-------------------------|--------------|-------------|--------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | (<i>p</i>) | Model ¹ | Model ² | |
| F _P -r | RC7 | SCL-90-R Anxiety | <0.001* | 0.033 | <0.001* | 0.03 | 0.10 | 0.08 |
| F_S | RC7 | SCL-90-R Anxiety | 0.003 | | | 0.03 | 0.07 | 0.05 |
| FBS-r | RC7 | SCL-90-R Anxiety | 0.023 | | | 0.03 | 0.06 | 0.03 |
| F-r | RC7 | SCL-90-R Phobic Anxiety | 0.181 | | | 0.01 | 0.01 | 0.02 |
| F _P -r | RC7 | SCL-90-R Phobic Anxiety | 0.128 | | | 0.01 | 0.01 | 0.02 |
| F_S | RC7 | SCL-90-R Phobic Anxiety | 0.263 | | | 0.01 | 0.01 | 0.01 |
| FBS-r | RC7 | SCL-90-R Phobic Anxiety | 0.183 | | | 0.01 | 0.01 | 0.02 |
| F-r | RC8 | SCL-90-R Psychoticism | <0.001* | <0.001*† | 0.478 | -0.01 | 0.18 | 0.19 |
| F _P -r | RC8 | SCL-90-R Psychoticism | <0.001* | <0.001*† | 0.182 | -0.01 | 0.09 | 0.11 |
| F_S | RC8 | SCL-90-R Psychoticism | <0.001* | <0.001*† | 0.370 | -0.01 | 0.09 | 0.10 |
| FBS-r | RC8 | SCL-90-R Psychoticism | < 0.001* | <0.001*† | 0.319 | -0.01 | 0.06 | 0.07 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|----------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (p) | (p) | (p) | Model ¹ | Model ² | |
| F-r | RC9 | EASI Impulsivity | 0.220 | | | 0.03 | 0.03 | 0.01 |
| F _P -r | RC9 | EASI Impulsivity | 0.069 | | | 0.03 | 0.04 | 0.02 |
| F_S | RC9 | EASI Impulsivity | 0.286 | | | 0.03 | 0.03 | 0.01 |
| FBS-r | RC9 | EASI Impulsivity | 0.331 | | | 0.03 | 0.03 | 0.01 |
| F-r | RC9 | Goldberg Mania Scale | 0.002* | 0.143 | 0.002* | 0.06 | 0.10 | 0.05 |
| F _P -r | RC9 | Goldberg Mania Scale | 0.002* | 0.107 | 0.001* | 0.06 | 0.10 | 0.05 |
| F_S | RC9 | Goldberg Mania Scale | 0.026 | | | 0.06 | 0.08 | 0.03 |
| FBS-r | RC9 | Goldberg Mania Scale | 0.013 | | | 0.06 | 0.09 | 0.04 |

Note. *p < .003 (Bonferroni corrected $\alpha = .05/15 = .003$). ¹Original model consists of the RC Scale as a predictor of the criterion. ²Final model includes the RC Scale, Validity Scale, and an interaction term (RC*Validity) as predictors of the criterion. [†]A visual inspection of the plotted regression lines indicated the significant findings were in the expected direction (Slope: the slope line for the association between the RC Scale and external criterion was steeper for those with lower validity scale scores; Intercept: When the raw RC Scale score = 0, the predicted criterion measure value was higher for those with low Validity Scale scores than for those with high Validity Scale scores, assuming a positive bivariate association between the RC Scale and criterion).

Significance Levels for Outpatient Sample Moderated Linear Regression Equations, Slope Differences, and Intercept Differences, and Associated R^2 Values (N's Range From 707 to 748)

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|--------------------|------------|--------------|--------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | (<i>p</i>) | (<i>p</i>) | Model ¹ | Model ² | |
| F-r | RCd | Feels overwhelmed | 0.027 | | | 0.09 | 0.10 | 0.01 |
| F _P -r | RCd | Feels overwhelmed | 0.556 | | | 0.09 | 0.09 | 0.00 |
| F_S | RCd | Feels overwhelmed | 0.357 | | | 0.09 | 0.09 | 0.00 |
| FBS-r | RCd | Feels overwhelmed | 0.010 | | | 0.09 | 0.10 | 0.01 |
| F-r | RCd | Tearful | 0.606 | | | 0.06 | 0.06 | 0.00 |
| F _P -r | RCd | Tearful | 0.122 | | | 0.06 | 0.06 | 0.01 |
| F_S | RCd | Tearful | 0.331 | | | 0.06 | 0.06 | 0.00 |
| FBS-r | RCd | Tearful | < 0.001* | 0.332 | < 0.001* | 0.06 | 0.10 | 0.04 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|-------------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| F-r | RCd | Feels gets raw deal from life | <0.001* | 0.001*† | 0.002* | 0.06 | 0.08 | 0.03 |
| F _P -r | RCd | Feels gets raw deal from life | <0.001* | 0.110 | < 0.001* | 0.06 | 0.08 | 0.02 |
| F_S | RCd | Feels gets raw deal from life | 0.084 | | | 0.06 | 0.06 | 0.01 |
| FBS-r | RCd | Feels gets raw deal from life | 0.142 | | | 0.06 | 0.06 | 0.01 |
| F-r | RCd | Sad | 0.006 | | | 0.11 | 0.12 | 0.01 |
| F _P -r | RCd | Sad | 0.117 | | | 0.11 | 0.11 | 0.01 |
| F_S | RCd | Sad | 0.162 | | | 0.11 | 0.11 | 0.00 |
| FBS-r | RCd | Sad | <0.001* | 0.009 | <0.001*† | 0.11 | 0.13 | 0.03 |
| F-r | RCd | Feels hopeless | 0.059 | | | 0.15 | 0.15 | 0.01 |
| F _P -r | RCd | Feels hopeless | 0.105 | | | 0.15 | 0.15 | 0.01 |
| F_S | RCd | Feels hopeless | 0.429 | | | 0.15 | 0.15 | 0.00 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|----------------------|--------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| FBS-r | RCd | Feels hopeless | 0.116 | | | 0.15 | 0.15 | 0.01 |
| F-r | RCd | Feels like a failure | 0.037 | | | 0.11 | 0.12 | 0.01 |
| F _P -r | RCd | Feels like a failure | 0.377 | | | 0.11 | 0.11 | 0.00 |
| F_S | RCd | Feels like a failure | 0.645 | | | 0.11 | 0.11 | 0.00 |
| FBS-r | RCd | Feels like a failure | 0.671 | | | 0.11 | 0.11 | 0.00 |
| F-r | RCd | Complains of fatigue | 0.006 | | | 0.05 | 0.06 | 0.01 |
| F _P -r | RCd | Complains of fatigue | 0.451 | | | 0.05 | 0.05 | 0.00 |
| F_S | RCd | Complains of fatigue | <0.001* | 0.023 | < 0.001* | 0.05 | 0.08 | 0.03 |
| FBS-r | RCd | Complains of fatigue | <0.001* | 0.029 | < 0.001* | 0.05 | 0.09 | 0.04 |
| F-r | RCd | Suicidal ideations | 0.001* | 0.314 | < 0.001* | 0.14 | 0.15 | 0.02 |
| F _P -r | RCd | Suicidal ideations | 0.001* | 0.935 | < 0.001* | 0.14 | 0.15 | 0.02 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|------------------------------------|--------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | (p) | Model ¹ | Model ² | |
| Fs | RCd | Suicidal ideations | 0.161 | | | 0.14 | 0.14 | 0.00 |
| FBS-r | RCd | Suicidal ideations | 0.004 | | | 0.14 | 0.15 | 0.01 |
| F-r | RC1 | Difficulty concentrating | 0.001* | 0.003 | 0.022 | 0.09 | 0.11 | 0.02 |
| F _P -r | RC1 | Difficulty concentrating | 0.010 | | | 0.09 | 0.10 | 0.01 |
| F_S | RC1 | Difficulty concentrating | 0.028 | | | 0.09 | 0.10 | 0.01 |
| FBS-r | RC1 | Difficulty concentrating | 0.070 | | | 0.09 | 0.10 | 0.01 |
| F-r | RC1 | Preoccupation with health problems | 0.066 | | | 0.18 | 0.19 | 0.01 |
| F _P -r | RC1 | Preoccupation with health problems | 0.131 | | | 0.18 | 0.19 | 0.01 |
| F_S | RC1 | Preoccupation with health problems | 0.294 | | | 0.18 | 0.18 | 0.00 |
| FBS-r | RC1 | Preoccupation with health problems | 0.051 | | | 0.18 | 0.19 | 0.01 |
| F-r | RC1 | Multiple somatic complaints | 0.013 | | | 0.17 | 0.18 | 0.01 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|---|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| F _P -r | RC1 | Multiple somatic complaints | 0.140 | | | 0.17 | 0.18 | 0.00 |
| F_S | RC1 | Multiple somatic complaints | 0.054 | | | 0.17 | 0.18 | 0.01 |
| FBS-r | RC1 | Multiple somatic complaints | 0.108 | | | 0.17 | 0.18 | 0.01 |
| F-r | RC1 | Hypochondriacal | 0.062 | | | 0.10 | 0.10 | 0.01 |
| F _P -r | RC1 | Hypochondriacal | 0.223 | | | 0.10 | 0.10 | 0.00 |
| Fs | RC1 | Hypochondriacal | 0.194 | | | 0.10 | 0.10 | 0.00 |
| FBS-r | RC1 | Hypochondriacal | 0.184 | | | 0.10 | 0.10 | 0.00 |
| F-r | RC1 | Physical symptoms in response to stress | 0.009 | | | 0.14 | 0.15 | 0.01 |
| F _P -r | RC1 | Physical symptoms in response to stress | 0.145 | | | 0.14 | 0.14 | 0.01 |
| Fs | RC1 | Physical symptoms in response to stress | 0.222 | | | 0.14 | 0.14 | 0.00 |
| FBS-r | RC1 | Physical symptoms in response to stress | 0.151 | | | 0.14 | 0.14 | 0.01 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|-------------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| F-r | RC2 | Tearful | 0.009 | | | 0.06 | 0.07 | 0.01 |
| F _P -r | RC2 | Tearful | 0.070 | | | 0.06 | 0.06 | 0.01 |
| F_S | RC2 | Tearful | 0.008 | | | 0.06 | 0.07 | 0.01 |
| FBS-r | RC2 | Tearful | <0.001* | 0.272 | < 0.001* | 0.06 | 0.10 | 0.04 |
| F-r | RC2 | Feels gets raw deal from life | <0.001* | 0.012 | < 0.001* | 0.06 | 0.08 | 0.03 |
| F _P -r | RC2 | Feels gets raw deal from life | <0.001* | 0.148 | < 0.001* | 0.06 | 0.08 | 0.03 |
| F_S | RC2 | Feels gets raw deal from life | 0.017 | | | 0.06 | 0.07 | 0.01 |
| FBS-r | RC2 | Feels gets raw deal from life | 0.024 | | | 0.06 | 0.07 | 0.01 |
| F-r | RC2 | Depressed | <0.001* | <0.001*† | 0.020 | 0.16 | 0.19 | 0.03 |
| F _P -r | RC2 | Depressed | 0.004 | | | 0.16 | 0.17 | 0.01 |
| Fs | RC2 | Depressed | < 0.001* | <0.001*† | 0.036 | 0.16 | 0.18 | 0.03 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|--------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| FBS-r | RC2 | Depressed | <0.001* | <0.001*† | <0.001* | 0.16 | 0.21 | 0.05 |
| F-r | RC2 | Sad | <0.001* | <0.001*† | 0.249 | 0.10 | 0.12 | 0.03 |
| F _P -r | RC2 | Sad | 0.010 | | | 0.10 | 0.11 | 0.01 |
| F_S | RC2 | Sad | <0.001* | <0.001*† | 0.153 | 0.10 | 0.12 | 0.02 |
| FBS-r | RC2 | Sad | <0.001* | <0.001*† | <0.001* | 0.10 | 0.14 | 0.04 |
| F-r | RC2 | Ruminates | <0.001* | <0.001*† | 0.024 | 0.04 | 0.07 | 0.03 |
| F _P -r | RC2 | Ruminates | <0.001* | 0.003 | 0.008 | 0.04 | 0.06 | 0.02 |
| F_S | RC2 | Ruminates | 0.001* | <0.001*† | 0.220 | 0.04 | 0.05 | 0.02 |
| FBS-r | RC2 | Ruminates | <0.001* | <0.001*† | 0.768 | 0.04 | 0.06 | 0.03 |
| F-r | RC2 | Suicidal ideations | < 0.001* | 0.041 | <0.001* | 0.12 | 0.16 | 0.04 |
| F _P -r | RC2 | Suicidal ideations | < 0.001* | 0.054 | < 0.001* | 0.12 | 0.15 | 0.03 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|--------------------|--------------|-------------|--------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | (<i>p</i>) | Model ¹ | Model ² | |
| Fs | RC2 | Suicidal ideations | <0.001* | 0.008 | 0.002* | 0.12 | 0.14 | 0.02 |
| FBS-r | RC2 | Suicidal ideations | 0.001* | 0.837 | < 0.001* | 0.12 | 0.14 | 0.02 |
| F-r | RC7 | Anxious | <0.001* | <0.001*† | 0.008 | 0.06 | 0.09 | 0.03 |
| F _P -r | RC7 | Anxious | <0.001* | <0.001*† | 0.100 | 0.06 | 0.08 | 0.03 |
| F_S | RC7 | Anxious | 0.010 | | | 0.06 | 0.07 | 0.01 |
| FBS-r | RC7 | Anxious | <0.001* | 0.002*† | 0.002* | 0.06 | 0.08 | 0.02 |
| F-r | RC7 | Worrier | <0.001* | <0.001*† | 0.382 | 0.06 | 0.08 | 0.02 |
| F _P -r | RC7 | Worrier | 0.008 | | | 0.06 | 0.07 | 0.01 |
| F_S | RC7 | Worrier | 0.001* | <0.001*† | 0.216 | 0.06 | 0.07 | 0.02 |
| FBS-r | RC7 | Worrier | <0.001* | <0.001*† | 0.001* | 0.06 | 0.09 | 0.04 |
| F-r | RC7 | Nervous | < 0.001* | <0.001*† | 0.017 | 0.05 | 0.07 | 0.03 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted | R^2 Value | ΔR^2 |
|-------------------|-------|--------------------|--------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | |
| | | | (<i>p</i>) | <i>(p)</i> | <i>(p)</i> | Model ¹ | Model ² | |
| F _P -r | RC7 | Nervous | 0.002* | 0.001*† | 0.270 | 0.05 | 0.06 | 0.02 |
| F_S | RC7 | Nervous | <0.001* | 0.024 | 0.001* | 0.05 | 0.06 | 0.02 |
| FBS-r | RC7 | Nervous | 0.001* | 0.048 | 0.001* | 0.05 | 0.06 | 0.02 |

Note. *p < .002 (Bonferroni corrected $\alpha = .05/22 = .002$). ¹Original model consists of the RC Scale as a predictor of the criterion. ²Final model includes the RC Scale, Validity Scale, and an interaction term (RC*Validity) as predictors of the criterion. [†]A visual inspection of the plotted regression lines indicated the significant findings were in the expected direction (Slope: the slope line for the association between the RC Scale and external criterion was steeper for those with lower validity scale scores; Intercept: When the raw RC Scale score = 0, the predicted criterion measure value was higher for those with low Validity Scale scores than for those with high Validity Scale scores, assuming a positive bivariate association between the RC Scale and criterion).

Significance Levels for Inpatient Sample Moderated Logistic Regression Equations, Slope Differences, and Intercept Differences, and Associated Nagelkerke Pseudo- R^2 Values (N's Range From 2,369 to 2,544)

| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelk | erke R^2 | ΔR^2 |
|-------------------|-------|-----------------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Val | lue | (Nagelkerke) |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Original | Final | |
| | | | | | | Model ¹ | Model ² | |
| F-r | RCd | Admit problem: Suicidal | <0.001* | <0.001* | <0.001*† | 0.08 | 0.10 | 0.02 |
| F _P -r | RCd | Admit problem: Suicidal | 0.030 | | | 0.08 | 0.09 | 0.01 |
| F_s | RCd | Admit problem: Suicidal | < 0.001* | 0.064 | <0.001*† | 0.08 | 0.11 | 0.03 |
| FBS-r | RCd | Admit problem: Suicidal | 0.118 | | | 0.08 | 0.08 | 0.00 |
| F-r | RCd | Admit medication: Antidepressants | 0.970 | | | 0.07 | 0.07 | 0.00 |
| F _P -r | RCd | Admit medication: Antidepressants | 0.009 | | | 0.07 | 0.07 | 0.00 |
| F_{S} | RCd | Admit medication: Antidepressants | 0.873 | | | 0.07 | 0.07 | 0.00 |
| FBS-r | RCd | Admit medication: Antidepressants | < 0.001* | 0.146 | <0.001*† | 0.07 | 0.08 | 0.01 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelke | erke R^2 | ΔR^2 |
|-------------------|-------|---------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Val | lue | (Nagelkerke) |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Original | Final | |
| | | | | | | Model ¹ | Model ² | |
| F-r | RCd | Helplessness/Hopelessness | <0.001* | < 0.001* | <0.001*† | 0.11 | 0.13 | 0.02 |
| F _P -r | RCd | Helplessness/Hopelessness | <0.001* | 0.540 | <0.001*† | 0.11 | 0.12 | 0.01 |
| F_{S} | RCd | Helplessness/Hopelessness | <0.001* | 0.786 | <0.001*† | 0.11 | 0.13 | 0.02 |
| FBS-r | RCd | Helplessness/Hopelessness | 0.008 | | | 0.11 | 0.12 | 0.01 |
| F-r | RCd | Worthlessness | 0.030 | | | 0.06 | 0.06 | 0.00 |
| F _P -r | RCd | Worthlessness | 0.107 | | | 0.06 | 0.06 | 0.00 |
| F_{S} | RCd | Worthlessness | <0.001* | 0.674 | < 0.001* | 0.06 | 0.07 | 0.01 |
| FBS-r | RCd | Worthlessness | 0.137 | | | 0.06 | 0.06 | 0.00 |
| F-r | RC1 | Chronic pain | 0.008 | | | 0.08 | 0.08 | 0.00 |
| F _P -r | RC1 | Chronic pain | <0.001* | 0.294 | <0.001*† | 0.08 | 0.09 | 0.01 |
| F_{S} | RC1 | Chronic pain | <0.001* | 0.346 | <0.001*† | 0.08 | 0.10 | 0.02 |
| FBS-r | RC1 | Chronic pain | 0.006 | | | 0.08 | 0.08 | 0.00 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelk | erke R^2 | ΔR^2 |
|-------------------|-------|-----------------------------------|------------|-------------|--------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Val | lue | (Nagelkerke) |
| | | | <i>(p)</i> | <i>(p)</i> | (<i>p</i>) | Original | Final | |
| | | | | | | Model ¹ | Model ² | |
| F-r | RC2 | Admit problem: Depression | 0.005 | | | 0.12 | 0.12 | 0.00 |
| F _P -r | RC2 | Admit problem: Depression | <0.001* | 0.545 | <0.001*† | 0.12 | 0.13 | 0.01 |
| F_{S} | RC2 | Admit problem: Depression | 0.001* | 0.118 | 0.001*† | 0.12 | 0.12 | 0.00 |
| FBS-r | RC2 | Admit problem: Depression | < 0.001* | <0.001*† | < 0.001* | 0.12 | 0.14 | 0.02 |
| F-r | RC2 | Admit medication: Antidepressants | 0.057 | | | 0.06 | 0.07 | 0.01 |
| F _P -r | RC2 | Admit medication: Antidepressants | 0.053 | | | 0.06 | 0.07 | 0.01 |
| F_{S} | RC2 | Admit medication: Antidepressants | 0.435 | | | 0.06 | 0.07 | 0.01 |
| FBS-r | RC2 | Admit medication: Antidepressants | < 0.001* | 0.126 | < 0.001* | 0.06 | 0.09 | 0.03 |
| F-r | RC2 | Mood: Depressed | 0.546 | | | 0.12 | 0.12 | 0.00 |
| F _P -r | RC2 | Mood: Depressed | 0.040 | | | 0.12 | 0.12 | 0.00 |
| F_S | RC2 | Mood: Depressed | 0.399 | | | 0.12 | 0.12 | 0.00 |
| FBS-r | RC2 | Mood: Depressed | < 0.001* | 0.589 | <0.001* | 0.12 | 0.13 | 0.01 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelke | erke R^2 | ΔR^2 |
|-------------------|-------|--------------------------|------------|-------------|-------------|--------------------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Val | ue | (Nagelkerke) |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Original | Final | |
| | | | | | | Model ¹ | Model ² | |
| F-r | RC2 | Loss of interest | 0.098 | | | 0.06 | 0.06 | 0.00 |
| F _P -r | RC2 | Loss of interest | 0.564 | | | 0.06 | 0.06 | 0.00 |
| F_S | RC2 | Loss of interest | 0.464 | | | 0.06 | 0.06 | 0.00 |
| FBS-r | RC2 | Loss of interest | <0.001* | 0.060 | < 0.001* | 0.06 | 0.07 | 0.01 |
| F-r | RC2 | Suicidal ideation | <0.001* | 0.837 | < 0.001* | 0.11 | 0.12 | 0.01 |
| F _P -r | RC2 | Suicidal ideation | 0.002* | 0.001* | 0.138 | 0.11 | 0.12 | 0.01 |
| F_{S} | RC2 | Suicidal ideation | <0.001* | < 0.001* | 0.374 | 0.11 | 0.12 | 0.01 |
| FBS-r | RC2 | Suicidal ideation | <0.001* | 0.022 | < 0.001* | 0.11 | 0.12 | 0.01 |
| F-r | RC6 | Admit problem: Psychoses | <0.001* | 0.358 | <0.001*† | 0.04 | 0.15 | 0.11 |
| F _P -r | RC6 | Admit problem: Psychoses | <0.001* | <0.001*† | <0.001*† | 0.04 | 0.10 | 0.06 |
| F_{S} | RC6 | Admit problem: Psychoses | <0.001* | <0.001*† | <0.001*† | 0.04 | 0.10 | 0.06 |
| FBS-r | RC6 | Admit problem: Psychoses | < 0.001* | <0.001*† | <0.001*† | 0.04 | 0.16 | 0.12 |

| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelk | erke R^2 | ΔR^2 |
|-------------------|-------|---------------------|------------|-------------|-------------|-----------|--------------------|--------------|
| Scale | Scale | | Moderation | Differences | Differences | Val | lue | (Nagelkerke) |
| | | | <i>(p)</i> | <i>(p)</i> | <i>(p)</i> | Original | Final | |
| | | | | | | $Model^1$ | Model ² | |
| F-r | RC6 | Paranoid/Suspicious | <0.001* | 0.093 | <0.001*† | 0.05 | 0.12 | 0.07 |
| F _P -r | RC6 | Paranoid/Suspicious | < 0.001* | <0.001*† | <0.001*† | 0.05 | 0.09 | 0.04 |
| F_{S} | RC6 | Paranoid/Suspicious | < 0.001* | <0.001*† | <0.001*† | 0.05 | 0.09 | 0.04 |
| FBS-r | RC6 | Paranoid/Suspicious | < 0.001* | <0.001*† | <0.001*† | 0.05 | 0.12 | 0.07 |
| F-r | RC6 | Delusions | < 0.001* | 0.561 | <0.001*† | 0.03 | 0.17 | 0.14 |
| F _P -r | RC6 | Delusions | < 0.001* | <0.001*† | <0.001*† | 0.03 | 0.11 | 0.08 |
| Fs | RC6 | Delusions | < 0.001* | <0.001*† | <0.001*† | 0.03 | 0.11 | 0.08 |
| FBS-r | RC6 | Delusions | < 0.001* | <0.001*† | <0.001*† | 0.03 | 0.16 | 0.13 |
| | | | | | | | | |

Note. *p < .004 (Bonferroni corrected $\alpha = .05/13 = .004$). ¹Original model consists of the RC Scale as a predictor of the log odds of the criterion. ²Final model includes the RC Scale, Validity Scale, and an interaction term (RC*Validity) as predictors of the log odds of the criterion. [†]A visual inspection of the plotted regression lines indicated the significant findings were in the expected direction (Slope: the slope line for the association between the RC Scale and the log odds of the criterion measure value was steeper for those with lower validity scale scores; Intercept: When the raw RC Scale score = 0, the predicted log odds of the criterion measure value was higher for those with low Validity Scale scores than for those with high Validity Scale scores, assuming a positive bivariate association between the RC Scale and criterion. The opposite pattern was predicted for those regressions with negative bivariate associations between the RC Scale and criterion).

Significance Levels for Forensic Sample Moderated Linear & Logistic Regression Equations, Slope Differences, and Intercept Differences, and Associated R^2 & Nagelkerke Pseudo- R^2 Values (N's Range From 1,138 to 1,358)

| | Linear Regression Results | | | | | | | | | |
|-------------------|---------------------------|------------------------------------|-----------------|--------------|--------------|--------------------|--------------------|--------------|--|--|
| Validity | RC | External Criterion | Overall | Slope | Intercept | Adjusted A | R^2 Value | ΔR^2 | | |
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | | | |
| | | | (<i>p</i>) | (<i>p</i>) | (<i>p</i>) | Model ¹ | Model ² | | | |
| F-r | RC1 | Previous medical problems (Summed) | < 0.001* | 0.002*† | <0.001*† | 0.04 | 0.08 | 0.04 | | |
| F _P -r | RC1 | Previous medical problems (Summed) | <0.001* | 0.112 | <0.001*† | 0.04 | 0.06 | 0.02 | | |
| F_{S} | RC1 | Previous medical problems (Summed) | 0.002* | 0.044 | 0.004*† | 0.04 | 0.05 | 0.01 | | |
| FBS-r | RC1 | Previous medical problems (Summed) | 0.053 | | | 0.04 | 0.05 | 0.01 | | |
| | | Log | istic Regressio | on Results | | | | | | |
| Validity | RC | External Criterion | Overall | Slope | Intercept | Nagelkerke | R^2 Value | ΔR^2 | | |
| Scale | Scale | | Moderation | Differences | Differences | Original | Final | (Nagelkerke) | | |
| | | | (<i>p</i>) | (<i>p</i>) | (<i>p</i>) | $Model^1$ | Model ² | | | |
| | | | | | | | | | | |

| F | -r RCd | Current medication: Antidepressants | 0.902 | | | 0.09 | 0.09 | 0.00 |
|----------------|---------|--------------------------------------|----------|-------|----------|------|------|------|
| F _P | -r RCd | Current medication: Antidepressants | 0.095 | | | 0.09 | 0.10 | 0.01 |
| F | s RCd | Current medication: Antidepressants | 0.893 | | | 0.09 | 0.09 | 0.00 |
| FB | S-r RCd | Current medication: Antidepressants | < 0.001* | 0.481 | <0.001* | 0.09 | 0.11 | 0.02 |
| F | -r RCd | Previous medication: Antidepressants | 0.401 | | | 0.09 | 0.09 | 0.00 |
| F _P | -r RCd | Previous medication: Antidepressants | 0.318 | | | 0.09 | 0.09 | 0.00 |
| F | s RCd | Previous medication: Antidepressants | 0.510 | | | 0.09 | 0.09 | 0.00 |
| FB | S-r RCd | Previous medication: Antidepressants | 0.002* | 0.822 | 0.001*† | 0.09 | 0.10 | 0.01 |
| F | -r RCd | Mood: Sad/Depressed | 0.594 | | | 0.09 | 0.09 | 0.00 |
| F _P | -r RCd | Mood: Sad/Depressed | 0.018 | | | 0.09 | 0.10 | 0.01 |
| F | s RCd | Mood: Sad/Depressed | 0.636 | | | 0.09 | 0.09 | 0.00 |
| FB | S-r RCd | Mood: Sad/Depressed | < 0.001* | 0.052 | <0.001* | 0.09 | 0.11 | 0.02 |
| F | -r RC2 | Current medication: Antidepressants | 0.680 | | | 0.11 | 0.11 | 0.00 |
| F _P | -r RC2 | Current medication: Antidepressants | 0.023 | | | 0.11 | 0.12 | 0.01 |
| F | s RC2 | Current medication: Antidepressants | 0.565 | | | 0.11 | 0.11 | 0.00 |
| FB | S-r RC2 | Current medication: Antidepressants | 0.001* | 0.539 | < 0.001* | 0.11 | 0.13 | 0.02 |

| F-r | RC2 | Previous medication: Antidepressants | 0.383 | | | 0.10 | 0.11 | 0.01 |
|---------------------------|-----|--------------------------------------|----------|----------|----------|------|------|------|
| F _P -r | RC2 | Previous medication: Antidepressants | 0.152 | | | 0.10 | 0.11 | 0.01 |
| F_{S} | RC2 | Previous medication: Antidepressants | 0.340 | | | 0.10 | 0.11 | 0.01 |
| FBS-r | RC2 | Previous medication: Antidepressants | 0.004* | 0.832 | 0.001* | 0.10 | 0.12 | 0.02 |
| F-r | RC2 | Current suicidal ideation | 0.004* | 0.042 | 0.008 | 0.20 | 0.22 | 0.02 |
| F _P -r | RC2 | Current suicidal ideation | 0.105 | | | 0.20 | 0.21 | 0.01 |
| $\mathbf{F}_{\mathbf{S}}$ | RC2 | Current suicidal ideation | 0.281 | | | 0.20 | 0.20 | 0.00 |
| FBS-r | RC2 | Current suicidal ideation | 0.288 | | | 0.20 | 0.20 | 0.00 |
| F-r | RC2 | Mood: Sad/Depressed | 0.053 | | | 0.12 | 0.12 | 0.00 |
| F _P -r | RC2 | Mood: Sad/Depressed | 0.016 | | | 0.12 | 0.13 | 0.01 |
| F_{S} | RC2 | Mood: Sad/Depressed | 0.146 | | | 0.12 | 0.12 | 0.00 |
| FBS-r | RC2 | Mood: Sad/Depressed | <0.001* | 0.007 | <0.001* | 0.12 | 0.14 | 0.02 |
| F-r | RC6 | Thought processes: Delusional | <0.001* | 0.056 | <0.001*† | 0.05 | 0.16 | 0.11 |
| F _P -r | RC6 | Thought processes: Delusional | < 0.001* | <0.001*† | 0.004*† | 0.05 | 0.14 | 0.09 |
| F_{S} | RC6 | Thought processes: Delusional | < 0.001* | <0.001*† | <0.001*† | 0.05 | 0.15 | 0.10 |
| FBS-r | RC6 | Thought processes: Delusional | <0.001* | <0.001*† | <0.001*† | 0.05 | 0.15 | 0.10 |
| | | | | | | | | |

Note. *p < .006 (Bonferroni corrected $\alpha = .05/9 = .006$). ¹Original model consists of the RC Scale as a predictor of the criterion (in the linear regressions) and of the log odds of the criterion (in the logistic regressions). ²Final model includes the RC Scale, Validity Scale, and an interaction term (RC*Validity) as predictors of the criterion/log odds of the criterion. ⁺A visual inspection of the plotted regression lines indicated the significant findings were in the expected direction (Slope: the slope line for the association between the RC Scale and the criterion/log odds of the criterion was steeper for those with lower validity scale scores; Intercept: When the raw RC Scale score = 0, the predicted criterion/log odds of the criterion value was higher for those with low Validity Scale scores than for those with high Validity Scale scores, assuming a positive bivariate association between the RC Scale and criterion. The opposite pattern was predicted for those regressions with negative bivariate associations between the RC Scale and criterion).

Chi Square Summary Analyses

In order to explore the patterns of significant findings across samples and Validity Scales, a series of chi square analyses were conducted. To ensure the reliability of chi square findings, it is recommended to avoid interpreting results with 20% or more expected cell counts less than five or any expected cell counts less than one (Yates, Moore, & McCabe, 1999). Unfortunately, because the forensic sample had so few moderated regressions to examine (due to the elimination of planned analyses because of weak bivariate correlational associations), the associated chi square analysis could not be meaningfully interpreted. The interpretable findings are presented below.

Differences in Overall Moderation Effects by Sample. The broadest chi square summaries involved examining whether there were differences in the frequency of overall moderation effects by sample (Table 12). Results indicate that there were significant differences in the frequency of moderation effects by sample, with the college sample having a higher frequency of statistically significant moderation effects than would be expected by chance, χ^2 (3, N = 88) = 14.43, p = .002. It should be noted, however, that the standardized residuals (+1.8 for the significant/college cell; -1.9 for the non-significant/college cell) were slightly lower than the recommended cut-off of |1.96| (indicating 95% confidence that the cell's findings are not due to chance; Field, 2010) for cells with practically significantly large discrepancies from what is expected by chance.

Differences in Overall Moderation Effects by Validity Scale. A chi square analysis examined whether there were differences in the frequencies of moderation effects by Validity Scale within each of the four samples. Table 13 indicates there were no significant differences in the frequency of overall moderation effects by Validity Scale within the college sample. Results were similar for the outpatient (Table 14) and inpatient (Table 15) samples, which also had non-significant overall moderation effects. The chi square results examining the forensic sample (Table 16) were not reliable enough to be interpreted.

Follow-Up Analyses

Although there were several significant moderation effects, there were also several nonsignificant findings. It is possible that these nonsignificant results were due to an inability of the Validity Scales to moderate criterion validity, but it is also possible that they were due to statistical artifacts related to violations of assumptions required for accurate optimal application of the methodology used to for moderation (Fairchild & MacKinnon, 2009; Field, 2009). Although we used a variety of samples analyzed in different assessment contexts and selected only conceptually and statistically associated RC Scale / Criterion Measure pairs, we also followed standard practices in the extant personality assessment literature, including not excluding moderation equations based upon issues such as multicollinearity of predictors, heteroskedasticity, independence of errors, non-normal distributions of errors, and nonlinearity. An advantage of this practice is that the sample analyzed is most representative of the population at large, and findings

| | Number of | Standardized | Number of | Standardized | χ^2 | р |
|--------------------|-------------|--------------|-------------|--------------|----------|-------|
| | Significant | Residual | Non- | Residual | | |
| | Findings | | Significant | | | |
| | | | Findings | | | |
| Overall Moderation | | | | | 10.860 | 0.013 |
| College | 38 | (+1.4) | 22 | (-1.4) | | |
| Outpatient | 36 | (-1.3) | 52 | (+1.3) | | |
| Inpatient | 31 | (+0.0) | 21 | (-0.9) | | |
| Forensic | 14 | (-1.0) | 22 | (+1.0) | | |

Chi Square Analyses Examining Frequency of Statistically Significant Overall Moderation Effects by Sample

Note. df for analyses = 3.

Chi Square Analyses Examining Frequency of Statistically Significant Overall Moderation Effects by Scale Within the College Sample

| | Number of | Standardized | Number of | Standardized | χ^2 | р |
|--------------------|-------------|--------------|-------------|--------------|----------|-------|
| | Significant | Residual | Non- | Residual | | |
| | Findings | | Significant | | | |
| | | | Findings | | | |
| Overall Moderation | | | | | 3.158 | 0.368 |
| F-r | 11 | (+0.5) | 4 | (-0.6) | | |
| F _P -r | 11 | (+0.5) | 4 | (-0.6) | | |
| Fs | 9 | (-0.2) | 6 | (+0.2) | | |
| FBS-r | 7 | (-0.8) | 8 | (+1.1) | | |
| | | | | | | |

Note. df for all analyses = 3.

Chi Square Analyses Examining Frequency of Statistically Significant Overall Moderation Effects by Scale Within the Outpatient Sample

| | Number of | Standardized | Number of | Standardized | χ^2 | р |
|--------------------|-------------|--------------|-------------|--------------|----------|-------|
| | Significant | Residual | Non- | Residual | | |
| | Findings | | Significant | | | |
| | | | Findings | | | |
| Overall Moderation | | | | | 3.009 | 0.390 |
| F-r | 11 | (+0.7) | 11 | (-0.6) | | |
| F _P -r | 7 | (-0.7) | 15 | (+0.6) | | |
| Fs | 7 | (-0.7) | 15 | (+0.6) | | |
| FBS-r | 11 | (+0.7) | 11 | (-0.6) | | |
| | | | | | | |

Note. df for all analyses = 3.

Chi Square Analyses Examining Frequency of Statistically Significant Overall Moderation Effects by Scale Within the Inpatient Sample

| | Number of | Standardized | Number of | Standardized | χ^2 | р |
|--------------------|-------------|--------------|-------------|--------------|----------|-------|
| | Significant | Residual | Non- | Residual | | |
| | Findings | | Significant | | | |
| | | | Findings | | | |
| Overall Moderation | | | | | 2.157 | 0.541 |
| F-r | 6 | (-0.6) | 7 | (+0.8) | | |
| F _P -r | 7 | (-0.3) | 6 | (+0.3) | | |
| Fs | 9 | (+0.4) | 4 | (-0.5) | | |
| FBS-r | 9 | (+0.4) | 4 | (-0.5) | | |
| | | | | | | |

Note. df for all analyses = 3.

Chi Square Analyses Examining Frequency of Statistically Significant Overall Moderation Effects by Scale Within the Forensic Sample

| | Number of | Standardized | Number of | Standardized | χ^2 | р |
|--------------------|-------------|--------------|-------------|--------------|----------|-------|
| | Significant | Residual | Non- | Residual | | |
| | Findings | | Significant | | | |
| | | | Findings | | | |
| Overall Moderation | | | | | 7.948† | 0.047 |
| F-r | 3 | (-0.3) | 6 | (+0.2) | | |
| F _P -r | 2 | (-0.8) | 7 | (+0.6) | | |
| F _S | 2 | (-0.8) | 7 | (+0.6) | | |
| FBS-r | 7 | (+1.9) | 2 | (-1.5) | | |
| | | | | | | |

Note. df for all analyses = 3; † denotes chi square analyses with 20% or greater expected cell counts less than 5, indicating the results are too unreliable to be meaningfully interpreted.

are thus more likely to generalize to the full range of individuals assessed in these settings. We also assumed that the external criterion measures were not impacted by selfreport bias because they involved mental health care professional ratings rather than selfreported problems (and in the college sample, self-reported scores before some participants were assigned to an overreporting group). However, upon examination of the results, it appears possible that some of our weak and nonsignificant findings might be due to contamination of the criteria by self-report or other statistical artifacts.

To explore these possibilities, we conducted some post-hoc examinations of the associations between some of our variables to determine whether we could identify any previously unexamined statistical explanations for weak and nonsignificant moderation findings. The exploratory findings detailed below provide some insight into factors that may have contributed to nonsignificant and weak moderating effects.

Independence of the Moderators (Validity Scales) and Dependent Variables (Criterion Measures). One concern we had regarding our power to detect existing moderation effects was whether the dependent variables were indeed independent of our moderators. Conceptually, an association between the moderator and dependent variable poses a major problem in this particular line of research. The samples and criteria were selected based upon the assumption that the criterion measures were not confounded by self-report. For example, we selected non-simulated samples with mental health professionals' ratings of examinee symptomatology, under the assumption that the expertise of the raters would lead them to detect and disregard fabricated self-reported symptoms while simultaneously observing and identifying genuine symptoms that were not reported by the examinee. Thus, our intention was for these ratings to be uncontaminated by the examinee's self-reported symptomatology. After an examination of the primary results of the study, where the Validity Scales operated as very strong moderators of criterion validity in the college sample but as weaker moderators in the three non-simulated samples, we questioned whether the ratings were actually independent of self-report bias. (Recall that independence of the criteria in the college sample was ensured by experimental manipulation).

To examine the independence of criteria from self-report bias, we calculated a series of correlations between overreporting Validity Scales and criterion measures in the four final samples (Tables 17 - 20). If the criterion measures were indeed independent of self-report biases, we would expect there to be no bivariate association between Validity Scale scores and ratings by independent professionals. Results for the college sample ranged from |.01| to |.28|, with the majority of correlations falling in the |.01| to |.10| range. Given the assignment of college students to their experimental group after completing their criterion measures, we can safely assume that their scores are not confounded by experimentally-assigned self-report bias. It is more likely that the strength of these correlations is in great part due to random variability in scores; therefore this level of association can serve as the baseline for evaluating the implications of correlations found in the remaining samples.

160

Bivariate Correlations Between Validity Scale Moderators and (a) Independent Variable Restructured Clinical (RC) Scales and (b) Dependent Variable External Criterion Measures in the College Sample (N's Range from 220 to 221)

| | F-r | F _P -r | Fs | FBS-r |
|------------------------------|-----|-------------------|-----|-------|
| Restructured Clinical Scales | | | | |
| RCd | .88 | .84 | .82 | .84 |
| RC1 | .88 | .86 | .93 | .92 |
| RC2 | .89 | .87 | .83 | .79 |
| RC3 | .67 | .67 | .63 | .47 |
| RC6 | .97 | .97 | .92 | .81 |
| RC7 | .85 | .82 | .81 | .80 |
| RC8 | .96 | .94 | .92 | .81 |
| RC9 | .58 | .59 | .58 | .43 |
| External Criterion Measures | | | | |
| SCL-90-R Depression | .09 | .06 | .08 | .20 |
| General Self Efficacy Scale | 05 | 03 | 02 | 13 |
| Rosenberg Self Esteem Scale | 10 | 06 | 09 | 21 |
| Cognitive Difficulties Scale | .10 | .06 | .07 | .14 |
| SCL-90-R Somatization | 04 | 09 | .01 | .16 |
| EASI Emotionality | .11 | .08 | .13 | .28 |

161

| EASI Sociability | 09 | 07 | 06 | 09 |
|----------------------------|-----|-----|-----|-----|
| SCL-90-R Hostility | .05 | .03 | .05 | .08 |
| SCL-90-R Paranoid Ideation | .11 | .08 | .11 | .13 |
| SCL-90-R Obsessive- | .07 | .05 | .07 | .14 |
| Compulsive | | | | |
| SCL-90-R Interpersonal | .09 | .06 | .07 | .17 |
| Sensitivity | | | | |
| SCL-90-R Anxiety | .06 | .02 | .06 | .15 |
| SCL-90-R Phobic Anxiety | .02 | .01 | .03 | .10 |
| SCL-90-R Psychoticism | .02 | 03 | 02 | .04 |
| EASI Activity | 13 | 12 | 09 | 11 |
| EASI Impulsivity | .05 | .02 | .05 | .03 |
| Goldberg Mania Scale | 02 | 02 | .03 | 05 |

Note. *p < .05; **p < .002 (Bonferroni corrected alpha = .05/25 = .002).

Bivariate Correlations Between Validity Scale Moderators and (a) Independent Variable Restructured Clinical (RC) Scales and (b) Dependent Variable External Criterion Measures in the Outpatient Sample (N's Range from 707 to 748)

| | F-r | F _P -r | Fs | FBS-r |
|------------------------------|-----|-------------------|-----|-------|
| Restructured Clinical Scales | | | | |
| RCd | .74 | .49 | .54 | .71 |
| RC1 | .74 | .48 | .78 | .85 |
| RC2 | .65 | .47 | .46 | .62 |
| RC3 | .55 | .45 | .42 | .25 |
| RC6 | .76 | .68 | .57 | .40 |
| RC7 | .77 | .55 | .60 | .63 |
| RC8 | .80 | .62 | .74 | .49 |
| RC9 | .37 | .32 | .28 | .09 |
| External Criterion Measures | | | | |
| Feels overwhelmed | .26 | .18 | .21 | .29 |
| Tearful | .16 | .06 | .17 | .31 |
| Feels gets raw deal from | .26 | .24 | .20 | .21 |
| life | | | | |
| Sad | .23 | .14 | .19 | .33 |
| Feels hopeless | .31 | .24 | .22 | .32 |

| Feels like a failure | .29 | .20 | .21 | .24 |
|---------------------------|-----|-----|-----|-----|
| Complains of fatigue | .21 | .13 | .25 | .30 |
| Suicidal ideations | .36 | .29 | .26 | .32 |
| Difficulty concentrating | .28 | .21 | .27 | .26 |
| Preoccupation with health | .29 | .20 | .30 | .34 |
| problems | | | | |
| Multiple somatic | .25 | .15 | .28 | .33 |
| complaints | | | | |
| Hypochondriacal | .18 | .09 | .21 | .23 |
| Physical symptoms in | .22 | .13 | .26 | .31 |
| response to stress | | | | |
| Depressed | .32 | .20 | .25 | .39 |
| Ruminates | .19 | .18 | .13 | .13 |
| Hostile | .14 | .18 | .07 | .03 |
| Low frustration tolerance | .24 | .25 | .19 | .11 |
| Cynical | .19 | .16 | .12 | .09 |
| Judgmental | .03 | .08 | .01 | 02 |
| Pessimistic | .30 | .25 | .22 | .23 |
| Critical of others | .08 | .13 | .07 | 02 |
| Indirect expression of | .09 | .07 | .08 | .04 |
| hostility | | | | |

| Difficulty trusting others | .16 | .17 | .09 | .06 |
|----------------------------|-----|-----|-----|-----|
| Delusional thinking | .13 | .17 | .09 | .01 |
| Suspicious | .15 | .18 | .08 | .03 |
| Psychotic symptoms | .14 | .16 | .06 | .01 |
| Paranoid features | .15 | .18 | .05 | .02 |
| Anxious | .25 | .18 | .21 | .24 |
| Fears losing control | .15 | .08 | .11 | .11 |
| Worrier | .21 | .15 | .18 | .25 |
| Irritable | .21 | .21 | .14 | .14 |
| Many specific fears | .22 | .17 | .20 | .18 |
| Nervous | .22 | .15 | .22 | .23 |
| Obsessive | .08 | .09 | .04 | .04 |
| Angry | .15 | .16 | .11 | .12 |
| Hallucinations | .17 | .18 | .12 | .04 |
| Energetic | 23 | 14 | 18 | 28 |
| Grandiose | 02 | .02 | 04 | 10 |
| Agitated | .26 | .23 | .21 | .19 |
| Accelerated speech | .04 | .06 | .06 | .02 |
| Excitable | .16 | .17 | .14 | .06 |
| Impulsive | .08 | .08 | .05 | 07 |

Note. *p < .05; **p < .002 (Bonferroni corrected alpha = .05/25 = .002).
Table 19

Point Biserial Correlations Between Validity Scale Moderators and (a) Independent Variable Restructured Clinical (RC) Scales and (b) Dependent Variable External Criterion Measures in the Inpatient Sample (N's Range from 2,369 to 2,544)

| | F-r | F _P -r | Fs | FBS-r |
|-------------------------------------|-----|-------------------|-----|-------|
| Restructured Clinical Scales | | | | |
| RCd | .77 | .48 | .53 | .67 |
| RC1 | .71 | .48 | .76 | .80 |
| RC2 | .65 | .40 | .37 | .58 |
| RC3 | .56 | .51 | .43 | .18 |
| RC6 | .73 | .73 | .55 | .34 |
| RC7 | .81 | .60 | .64 | .58 |
| RC8 | .78 | .66 | .74 | .44 |
| RC9 | .44 | .40 | .43 | .12 |
| External Criterion Measures | | | | |
| Admit problem: Suicidal | .15 | .08 | .02 | .15 |
| Admit medication: Antidepressants | .16 | .06 | .10 | .22 |
| Helplessness/Hopelessness | .17 | .07 | .05 | .20 |
| Worthlessness | .12 | .06 | .03 | .11 |
| Discouraged | .02 | .03 | .02 | .03 |
| Stressors: Chronic medical problems | .05 | 02 | .07 | .11 |

| Stressors: Acute medical problems | .02 | .01 | .04 | .04 |
|-----------------------------------|-----|-----|-----|-----|
| Chronic pain | .12 | .03 | .09 | .18 |
| Memory problems | .13 | .08 | .13 | .14 |
| History of head injury | .06 | .04 | .04 | .03 |
| Admit problem: Depression | .15 | .02 | .05 | .25 |
| Mood: Depressed | .20 | .10 | .11 | .25 |
| Loss of interest | .14 | .06 | .09 | .19 |
| Anhedonia | .09 | .02 | .04 | .15 |
| Suicidal ideation | .25 | .14 | .09 | .23 |
| Suicidal attempt | .18 | .11 | .07 | .13 |
| Suicidal plan | .15 | .09 | .03 | .15 |
| Ruminations | .02 | .00 | .02 | .04 |
| Admit problem: Psychoses | 07 | .04 | 01 | 19 |
| Admit medication: Antipsychotics | .08 | .09 | .06 | .01 |
| Paranoid/Suspicious | 02 | .05 | 02 | 13 |
| Ideas of reference | .06 | .08 | .04 | 02 |
| Delusions of reference | .00 | .03 | .01 | 03 |
| Delusions | 10 | .01 | 04 | 18 |
| Admit problem: Anxiety | .00 | 02 | .04 | .08 |
| Admit problem: Increase in PTSD | .16 | .09 | .18 | .16 |
| symptoms | | | | |

| Admit medication: Anxiolytics | .08 | .03 | .10 | .12 |
|-------------------------------|-----|-----|-----|-----|
| Mood: Anxious | .04 | .01 | .04 | .07 |
| Mood: Angry/Irritable | .00 | .02 | .03 | 01 |
| Worry | .01 | 01 | .00 | .02 |
| Panic | .09 | .04 | .06 | .11 |
| Compulsions | 01 | 02 | 03 | .00 |
| Obsessions | 03 | 01 | 03 | 03 |
| Nightmares | .22 | .16 | .22 | .23 |
| Flashbacks | .20 | .13 | .20 | .19 |
| Intrusive thoughts | .11 | .09 | .12 | .10 |
| Racing thoughts | .07 | .06 | .08 | .00 |
| Disorganized | 09 | .00 | 03 | 13 |
| Admit medication: Antimanics | .06 | .04 | .05 | .02 |
| Impulsive | 01 | .00 | 04 | 03 |

Note. *p < .05; **p < .002 (Bonferroni corrected alpha = .05/25 = .002).

Table 20

Bivariate and Point Biserial Correlations Between Validity Scale Moderators and (a) Independent Variable Restructured Clinical (RC) Scales and (b) Dependent Variable External Criterion Measures in the Forensic Sample (N's Range from 1,138 to 1,358)

| | F-r | F _P -r | F _S | FBS-r |
|---|-----|-------------------|----------------|-------|
| Restructured Clinical Scales | | | | |
| RCd | .82 | .59 | .66 | .73 |
| RC1 | .76 | .57 | .78 | .84 |
| RC2 | .70 | .53 | .54 | .64 |
| RC3 | .52 | .45 | .44 | .21 |
| RC6 | .83 | .78 | .65 | .54 |
| RC7 | .82 | .66 | .70 | .66 |
| RC8 | .85 | .71 | .78 | .60 |
| RC9 | .44 | .36 | .40 | .16 |
| External Criterion Measures | | | | |
| Previous medical problems (Summed) | .06 | .02 | .12 | .18 |
| Current medication: Antidepressants† | .21 | .12 | .17 | .25 |
| Previous medication: Antidepressants [†] | .22 | .15 | .17 | .24 |
| Mood: Sad/Depressed† | .20 | .09 | .17 | .26 |
| Current suicidal ideation [†] | 26 | 17 | 16 | 18 |
| Thought processes: Delusional [*] | .02 | .07 | .00 | 02 |

Note. *p < .05; **p < .002 (Bonferroni corrected alpha = .05/25 = .002). †Point biserial correlation with a binomial criterion.

Outpatient correlations ranged from |.01| to |.39|, with most falling in the |.10| to |.25| range. Results for the inpatient sample ranged from |<.01| to |.28|, with most point biserial correlations falling between |.05| and |.20|. Finally, forensic sample bivariate and point biserial correlations ranged from |<.01| to |.26|, with scores falling fairly evenly throughout the continuum of scores. These findings indicate that, in all four samples, at least some criteria were notably associated with the Validity Scales. Correlations tended to be highest in the outpatient sample, suggesting that the ratings in this sample were most likely to be confounded by self-report bias. This is not entirely surprising, given that mental health professionals in an outpatient setting tend to rely to a great degree on patient self-reported symptoms. Albeit weaker, there is evidence that ratings in the inpatient and forensic samples were somewhat confounded by self-report bias as well.

Multicollinearity Between the Independent Variables (RC Scales) and Moderator Variables (Validity Scales). Another concern we had regarding our power to detect existing moderation effects was whether our independent variables were too highly correlated with our moderators, thus leading to concerns about multicollinearity among predictors. Jaccard and Turrisi (2003) alleviate concerns about high multicollinearity between predictors and the interaction variable (e.g., RC1 correlated with an [RC1*F-r] variable or F-r correlated with an [RC1*F-r] variable), stating that even high intercorrelations among these variables are not problematic unless they are so correlated as to cause errors in the statistical package's ability to run the analysis. For instance, perfect multicollinearity, also called singularity, prevents the model from running (Cohen, Cohen, West, & Aiken, 2003; Jaccard & Turrisi, 2003). In the current study, none of the analyses led to singularity.

Despite there being no concerns regarding high intercorrelations between firstorder and second-order interaction terms, Jaccard and Turrisi (2003), warn that high multicollinearity between the first-order variables themselves (e.g., the RC1 variable correlated with the F-r variable) can indeed lead to serious problems. Specifically, predictor variable correlations above approximately .80 to .90 (Field, 2009) may lead to a variety of problems including inaccurate estimates of individual regression coefficients and large standard errors (Cohen, Cohen, West, & Aiken, 2003). In order to determine whether there were highly multicollinear independent variable and moderator predictors in the current study, we examined the pattern of bivariate correlations between the overreporting Validity Scales and the RC Scales across the four samples (Tables 17 – 20). The college sample had the highest frequency of high intercorrelations, with 23 of the 32 correlations at or above |.80| and one additional correlation in the |.70| to |.79|range (total range: .43 to .97). The outpatient sample had two of 32 correlations at or above |.80| and another seven in the |.70| to |.79| range (total range: .09 to .85). The inpatient sample had two of 32 correlations at or above |.80| and another six in the |.70| to

|.79| range (total range: .12 to .81). The forensic sample had five of 32 correlations at or above |.80| and another eight in the |.70| to |.79| range (total range: .16 to .85).

While examining the Validity Scale / RC Scale intercorrelations (Tables 17 - 20) to assess for multicollinearity, we also examined the pattern of these correlations to determine whether there was evidence that the Validity Scales were operating in an expected manner. The pattern of Validity Scale / RC Scale bivariate correlations was quite similar across samples, with the exception that the strength of the correlations was greater in the college sample. For all samples, the correlations followed a pattern consistent with our understanding of the utility of the Validity Scales. F-r, which is broad in scope and detects overreported psychopathology, was highly associated with the RC Scales associated with depression, somatic complaints, and psychotic symptoms (RCd, RC1, RC2, RC6, RC7, and RC8). F_{P} -r, which detects exaggerated rare symptoms, was most associated with RC6 and RC8. F_s and FBS-r, designed to detect overreported somatic and cognitive problems, respectively, were both most associated with RC1. Interestingly, F_8 was also consistently associated with RC8 across samples. These findings indicate that the Validity Scales were generally associated with the RC Scales in an expected manner.

Discussion

McGrath, Mitchell, Kim, and Hough (2010) questioned whether there is empirical support for the ability of existing validity indices to moderate substantive scale criterion validity. Their review found very limited support for the moderating abilities of extant validity scales. However, the studies available and selected for their review had several limitations (e.g., methodological concerns, use of weak external criteria, examination of outdated validity indices, utilization of samples with no motivation to distort their responses) which make it difficult to fairly examine the ability of validity scales to moderate substantive scale criterion validity. The current study aimed to examine whether the overreporting Validity Scales of a psychometrically supported self-report inventory (MMPI-2-RF) could moderate the criterion validity of several of the instrument's substantive scales (Research Question One). Further, we aimed to examine whether patterns of significant findings differed across settings where participants were believed to have varying levels of motivation to provide inaccurate responses (Research Question Two).

In order to avoid methodological limitations found in the current literature, four large archival samples were selected that had MMPI-2-RF data and a variety of conceptually-relevant external criteria which we identified as being uncontaminated by self-report biases. The samples were also selected because we believed they were likely to include participants with at least some (and up to a great deal of) motivation to exaggerate their genuine psychological problems, so that we could avoid null findings due to range restriction of scores.

Validity Scales as Moderators of Substantive Scale Criterion Validity

Regression analyses were conducted to examine the ability of MMPI-2-RF overreporting Validity Scales to moderate conceptually- and statistically-associated RC Scale/Criterion Measure pairs. Further, we examined whether significant moderations were due to slope differences, intercept differences, or both. Additionally, we examined whether the significant slope and intercept findings were in the hypothesized direction and the practical strength of significant findings. Finally, we conducted a series of chi square analyses to determine whether there were differences in the pattern of overall moderation results across samples and Validity Scales.

Overall, several of the moderated regression analyses indicated that the overreporting Validity Scales moderated RC Scale / Criterion Measure associations. Significant moderations occurred most frequently in the college and inpatient samples. An examination of practical significance indices demonstrated that the moderators explained the most additional variance in the college sample in statistically significant models (with 12.11% additional variance explained) as compared to the other samples (with additional variance explained ranging from 2.81% to 4.29%). This differentially larger impact of moderators in the college sample is not surprising, as the simulation design utilized in that sample helped control factors that likely led to more measurement error and noise in the three non-simulation samples, providing greater internal validity.

174

Further, when individuals are asked to exaggerate problems in simulation studies, they tend to report more extreme scores than occur in non-simulation samples. Although this tendency leads to weaker external validity in simulation samples, it also tends to lead to stronger results statistically due to greater distinctions between scores of exaggerators and controls (Rogers, 2008c). In sum, the overall findings from the college sample provide evidence for the Validity Scales as moderators under ideal conditions, whereas the outpatient, inpatient, and forensic samples provide evidence for how the Validity Scales operate as moderators under more ecologically valid conditions.

Moderation Due to Slope and Intercept Differences

Although specific hypotheses were not developed regarding whether significant moderating effects would be due to slope differences, intercept differences, or both, the pattern of significant findings varied, with some due to each individually and others due to combinations of both. Slope differences were most common in the college sample, occurring in 76% of significant college sample moderations and always in the expected direction. Although occurring less frequently, the significant slope effects also tended to be in the expected direction in the three non-simulated samples. These findings provide evidence for the idea that the MMPI-2-RF overreporting Validity Scales moderate RC Scale criterion validity in an expected manner, with higher Validity Scale scores leading to poorer RC Scale criterion validity. In other words, elevations on the Validity Scales indicate that the RC Scales are less associated with genuine symptomatology.

Significant intercept differences were most common in the inpatient and forensic samples, occurring in around 93% of the significant moderation equations in those samples. Results were mixed, however, regarding whether findings were in the expected direction. Figure 7 graphically depicts the pattern of several findings, where there were significant slope differences in the expected direction but also significant intercept differences in the unexpected direction. Given the frequency of such results, perhaps our expectation for the direction of intercept differences was unrealistic. It may be more realistic to consider the utility of Validity Scale scores at high and low RC Scale scores separately. For instance, at higher levels of RC Scale scores, the Validity Scales perform as anticipated, with higher Validity Scale scores indicating decreased criterion validity (e.g., where higher RC Scale scores not necessarily being linked to high criterion scores). However, at lower RC Scale scores, exaggerators are deemed by evaluators to be having



Figure 7. Interpreting Significant Intercept Differences in the Unanticipated Direction (In the Presence of Anticipated Significant Slope Differences)

greater symptoms than are honest responders. Findings at this end of the spectrum may be multivariate extrapolations that should not be interpreted, as discussed below.

The Trade-off between Generalizability and Statistical Accuracy. Our unanticipated intercept findings may be an artifact of the methods used to conduct the regression analyses. For instance, it is very unlikely to see individuals with high overreporting Validity Scale scores (exaggerators) who also have very low RC Scale scores, as the act of exaggeration generally increases their RC Scale scores. Unfortunately, this means that there are likely to be some multivariate outliers (with low Validity Scale and RC Scale scores) in the datasets used to conduct the regression analyses. Standard protocol in MMPI research is to avoid excluding cases for statistical purposes whenever possible, so that the results of the analyses can best generalize to test results of future examinees. In contrast, in ideal statistical practice, both univariate and multivariate outliers are removed from analyses in order to avoid significant influence from outlying data points. These two methodological philosophies, with different goals in mind, may lead to different results and conclusions. Future investigations should explore this possibility by applying more rigorous subject exclusion rules than is the practice in MMPI research.

In sum, given that it is rare to find low Validity Scale / RC Scale combinations, it is safest to limit the interpretation of our findings to the higher end of the RC Scale spectrum (e.g., the right side of Figure 7) in order to avoid extrapolation of our findings. Thus, although several of the intercept differences were in the unanticipated direction, even those were almost always in the expected direction if we limited our examination to the higher end of the spectrum. A consideration of the impact of multivariate outliers on the regression results will be discussed later.

Support for Each Individual Validity Scale as a Moderator of Criterion Validity

Overall, we found that all four of the overreporting Validity Scales served as statistically and practically significant moderators of RC Scale criterion validity, due to both slope and intercept effects. Further, we examined the pattern of significant moderation effects within each sample to determine how each scale performed in different simulation and non-simulated conditions. In sum, F-r and F_{P} -r were the most frequent moderators in the college sample. This is not surprising, given that half of the college sample was instructed to exaggerate severe psychopathology (as detected by F-r and F_{P} -r) but not somatic complaints (as detected by F_{S} and FBS-r).

In the outpatient sample, chi square summary analyses indicated there was no statistically meaningful difference in the frequency of significant moderations by Validity Scale. Although the inpatient chi square analysis was also nonsignificant, an examination of the results suggests that F_S and FBS-r moderated criterion validity somewhat more frequently (around 69% of the time) than did F-r and F_P -r (around 47% to 54% of the time). This is somewhat surprising, but perhaps reflects that there may be some somatic exaggerating in an inpatient setting in order to seek prolonged hospitalization or pain medications. Alternatively, given that the chi square findings were nonsignificant, this pattern may be artifactual.

Surprisingly, FBS-r was the most frequent moderator (78%) in the forensic sample, as compared to the other Validity Scales (which ranged from 22% to 33%). This is contrary to what would be expected, since forensic examinees historically tend to exaggerate psychiatric problems rather than somatic ones. However, the number of analyses was so small that the results of the chi square analysis were not interpretable. Thus, this result may be artifactual.

In sum, we were able to find that the MMPI-2-RF Validity Scales frequently moderated RC Scale criterion validity, although the patterns of significance (e.g., due to slope versus intercept differences, between samples, and between Validity Scales) were mixed. Further, the practical significance of the findings was much stronger in the college simulation sample than in the three non-simulated samples.

Considerations from the Follow-Up Analyses

In considering the results of our analyses we were concerned that the findings may have been weakened by the performance of analyses using data that did not meet conceptual expectations (e.g., independence of the criterion measures from self-report bias) or statistical assumptions (e.g., multicollinearity among predictors). We therefore conducted post-hoc analyses of the associations among variables to learn more about whether such factors may have impacted the results. Findings from these analyses provided some evidence that our criteria were more confounded by self-report bias than we originally assumed. Our post-hoc examination of the associations between the Validity Scales and criterion measures helped us to examine whether there was a

179

problematic level of multicollinearity among predictors. We found that there was a substantial frequency and strength of the Validity Scale / RC Scale associations, especially in the college sample. Interestingly, despite these high intercorrelations (which could overestimate standard error and thus lead to nonsignificance), the college sample had the strongest moderated regression findings of all of the samples. It is likely that the level of multicollinearity was not high enough to pose a problem, though it may have artifactually attenuated the obtained effect sizes. To be cautious, it may be more prudent to avoid interpreting regression analyses in future studies when the bivariate correlations between the validity scale and substantive scale are too high (e.g., above .80 or .90).

Implications of the Study Findings

This investigation provided evidence for the ability of MMPI-2-RF overreporting Validity Scales to moderate RC Scale criterion validity across four diverse samples. Results were especially strong in the college simulation sample, indicating that these scales function as moderators very well under ideal, controlled conditions. This is especially important given that there are a variety of factors (e.g., unreliable or invalid criterion measures, difficulties meeting the statistical assumptions) that can lead to difficulties in obtaining the statistical power to detect true moderating effects. However, such simulation samples cannot provide direct evidence of how well these scales moderate in more ecologically valid samples. Although the findings were weaker, we still found support for all four of the Validity Scales as moderators in outpatient, inpatient, and forensic samples. The current study also provided some insight into how slope and intercept differences can operate in the context of Validity Scale moderation of criterion validity. Typically, researchers tend to focus on slope effects, not paying much attention to how intercept differences play a role in moderation. Our results included both, and the findings did not strictly adhere to a pattern of slope-only or intercept-only findings across samples. Thus, researchers need to be careful to look for *both* slope and intercept differences when examining the moderating abilities of a Validity Scale. Furthermore, we found that nearly all of the significant slope differences were in the expected direction, indicating that when the overreporting Validity Scales moderate criterion validity, they consistently indicate that higher Validity Scale scores are associated with poorer RC Scale criterion validity. These are promising findings, and support the hypothesis that these scales are useful in detecting overreported symptoms.

Our findings also provided novel information about the meaning of intercept differences in Validity Scale moderation analyses. A large number of the significant intercept differences were *not* in the expected direction. There was more commonly an X-shaped pattern in the plotted regression lines (see Figure 7). This suggests that we may have had inaccurate expectations of how an intercept difference operates in this context. The findings suggest that the Validity Scale acts as a moderator as expected at high levels of RC Scale & Validity Scale scores but perhaps less useful for interpretation at lower scores. This is consistent with current use of these scales, as we do not tend to put a great deal of interpretive stock into the meaning of Validity Scales at lower scores.

Clinical Implications. An extensive literature examines the construct validity, criterion validity, test-retest reliability, positive and negative predictive power, and other psychometric properties of the MMPI-2 and MMPI-2-RF Validity Scales. This extant literature supports the utility of these scales in identifying invalid test protocols. Such empirical support is crucial for practicing clinical, assessment, and forensic psychologists who rely upon the results of these tests to make life-impacting clinical decisions. The current study extends our understanding of the utility of four of the MMPI-2-RF overreporting Validity Scales, demonstrating that they indeed moderate RC Scale criterion validity. Clinically, this indicates that the scales effectively detect when RC Scale responses inaccurately reflect genuine psychopathological symptoms and traits. In other words, when the overreporting Validity Scales are elevated, scores on the RC Scales are less likely to accurately reflect the examinee's psychological problems. Thus, our findings allow researchers and clinicians to have confidence in the ability of the examined scales to serve their intended purpose.

Although future research is needed to replicate these findings and examine the positive and negative predictive power of these scales at the recommended cut scores, the current study calls into question conclusions made by McGrath and colleagues (2010) who indicated that there was little support for the use of extant validity indices. At the time of that study, no MMPI-2-RF criterion validity moderation studies were available for review. Thus, these findings provide necessary support for the utility of the relatively new MMPI-2-RF Validity Scales across several samples including in outpatient,

182

inpatient, and forensic assessment settings, with elevations on these scales indicating that the substantive scales of the instrument are likely inaccurate portrayals of examinee psychological problems.

Limitations & Future Directions

The current study examined the utility of the MMPI-2-RF overreporting scales as moderators of RC Scale criterion validity. It would be informative to examine whether there is support for the moderating abilities of MMPI-2-RF Validity Scales examining non-content-based invalid responding (e.g., VRIN-r, TRIN-r) and overreporting (e.g., Lr, K-r), as well as the new overreporting Validity Scale, RBS. It would also be useful to examine the validity indices of other psychometrically-sound multiaxial personality inventories (e.g., MMPI-2, PAI, MCMI-III), as McGrath and colleagues (2010) reviewed findings from several outdated indices. Results utilizing current instruments and their validity indices are relevant for our understanding of the utility of contemporary validity indices.

Further, these analyses should be replicated in samples where there is a full range of non-content-based invalid responding (e.g., standard college, severe psychopathology, normative with no motivation to pay attention), underreporting (surgery candidate, police or other employment candidate, parenting evaluation), and overreporting (forensic, inpatient), respectively. This is especially important because McGrath's review made conclusions based upon nonsignificant moderation effects from samples where participants had no motivation to distort their responses. Because we would expect such samples to have range restriction in scores and thus to null findings, researchers should be thoughtful in the selection of their samples for these analyses.

In the current study, we discussed the dilemma that researchers face when attempting to strike a balance between obtaining statistically sound conclusions (utilizing "clean" data which meet the assumptions of the test) and ecologically meaningful conclusions (utilizing original generalizable samples). Future researchers should be aware of this dilemma and make informed decisions regarding how liberally or conservatively to adhere to each assumption of the chosen analyses and what impact their decisions may have on their findings. Table 21 includes recommendations for "best practices" for future researchers to consider when researching the ability of validity indices to moderate substantive scale criterion validity, including a recommendation to explore the impact of statistical decisions on the findings. Several factors in a moderated regression analysis can lead to nonsignificant or otherwise inaccurate findings, regardless of the actual utility of the validity scale being examined. Only after a systematic examination of the other factors (discussed above) and whether they contribute to nonsignificance should we conclude that the validity scale is an ineffective moderator. Thus, it is imperative that future researchers examine and discuss the psychometric properties of their data before making sweeping conclusions about the effectiveness of the validity scales as moderators.

Table 21

Recommendations for Best Practice in Examining Whether Validity Scales Moderate Substantive Scale Criterion Validity

| Issue | Recommendation |
|-----------------|---|
| Sample | • To avoid range restriction, select samples which include a variety |
| Selection | of valid and invalid responders (specific to the type of invalid |
| | responding your validity scales measure). |
| Validity Scale | • This line of research is so new that we currently need to examine all |
| Selection | extant validity indices, including those from personality |
| | inventories, stand-alone scales which detect exaggeration, and |
| | cognitive effort tests. |
| Substantive | • Ideally, we should examine <i>all</i> of the substantive scales of extant |
| Scale Selection | personality inventories. Additionally, it would also be useful to |
| | examine multi-test combinations (e.g., the ability of the TOMM to |
| | moderate the WAIS-IV Full Scale IQ and Subtests). |
| Criterion | • Criterion measures should be conceptually and statistically |
| Measure | associated with the substantive scales. They should not be |
| Selection | confounded by self-report bias. |
| Conducting | • Hierarchical Moderated Linear Regression: Use with ratio, interval, |
| Moderated | and (less ideally) ordinal criterion measures. |
| Regression | • Hierarchical Moderated Logistic Regression: Use with binomial |

| Analyses | criterion measures. |
|---------------|--|
| | • See Appendix A for Sample SPSS Syntax |
| | • Examine overall moderation, slope effects, and intercept |
| | effects separately. |
| | • Use a full model with all predictors when calculating beta |
| | weights used in graphical depictions. |
| | • Graphing the findings allows for an examination of whether |
| | significant slope and intercept differences occurred in the |
| | expected direction. |
| Examining the | • Look for and consider removing univariate and multivariate |
| Assumptions | outliers as well as influential cases which may affect the |
| of Linear & | regression line. |
| Logistic | • Examine the multivariate distribution of scores to avoid |
| Regression | extrapolating findings to levels of scores that were not |
| | present in the sample analyzed. |
| | • Large enough sample size: Power analyses (e.g., in the |
| | G*Power statistical package) can indicate appropriate sample |
| | sizes. |
| | Appropriate variable type: |
| | • Linear regression requires a continuous-level |
| | dependent variable; logistic regression requires a |

bivariate dependent variable.

- Bivariate, dummy, and continuous-level independent and moderator variables can be used.
- No multicollinearity:
 - Bivariate correlations above .80-.90 may be problematic.
 - VIF and Tolerance values may be useful in determining whether multicollinearity is present.
- Homoskedasticity: The variance of the residuals should be constant at all levels of the predictor variables. Standardized residuals plotted against standardized predicted values help determine whether this assumption is violated.
- Normally distributed errors: Examine histograms and normal probability plots to determine whether the errors are normally distributed.
- Independence of Errors: For any two data points, the residuals should be uncorrelated. Test for independence with the Durbin-Watson test.
- Linearity:
 - In linear regression, each predictor should have a linear relationship with the dependent variable, as evidenced

by scatterplots. In logistic regression, each predictor should have a linear relationship with the *natural log of* the dependent variable.

• The interested researcher is referred to Field (2009) for further details.

Appendix A

Example SPSS Syntax for Running Hierarchical Moderated Linear and Logistic

Regressions Using rfRCd (Substantive Scale), rfF (Validity Scale) Predicting Criteria

*Linear Regression: rfF moderating rfRCd*DEPraw *STEP 1 --- IV // Mod + IV*Mod REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT DEPraw /METHOD=ENTER rfRCd /METHOD=ENTER rfF FRCd . Execute. *STEP 2 --- IV + Mod // IV*Mod REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT DEPraw /METHOD=ENTER rfRCd rfF /METHOD=ENTER FRCd . Execute. *STEP 3 --- IV // Mod

REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT DEPraw /METHOD=ENTER rfRCd /METHOD=ENTER rfF . Execute. *STEP 4 --- Getting Full Regression Equation for Graphs REGRESSION /MISSING LISTWISE /STATISTICS COEFF OUTS R ANOVA CHANGE /CRITERIA=PIN(.05) POUT(.10) /NOORIGIN /DEPENDENT DEPraw /METHOD=ENTER rfRCd rfF FRCd . Execute.

*STEP 1 --- IV // Mod + IV*Mod LOGISTIC REGRESSION VARIABLES apsuic /METHOD=ENTER rfRCd /METHOD=ENTER rfF FRCd /PRINT = GOODFIT CI(95) /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) . EXECUTE.

*STEP 2 --- IV + Mod // IV*Mod LOGISTIC REGRESSION VARIABLES apsuic /METHOD=ENTER rfRCd rfF /METHOD=ENTER FRCd /PRINT = GOODFIT CI(95) /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) . EXECUTE.

*STEP 3 --- IV // Mod

LOGISTIC REGRESSION VARIABLES apsuic /METHOD=ENTER rfRCd /METHOD=ENTER rfF /PRINT = GOODFIT CI(95) /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) . EXECUTE.

*STEP 4 --- Getting Full Regression Equation for Graphs LOGISTIC REGRESSION VARIABLES apsuic /METHOD=ENTER rfRCd rfF FRCd /PRINT = GOODFIT CI(95) /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) . EXECUTE.

References

- Aday, L. A., Cliu, G. Y., & Anderson, R. (1980). Methodological issues in health care surveys of the Spanish-descent population. *American Journal of Public Health*, 70, 367-374.
- Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment screening. *Journal of Applied Psychology*, *95*, 648-680.
- Amador, X. F., Strauss, D. H., Yale, S. A., & Gorman, J. M. (1991). Awareness of illness in schizophrenia. *Schizophrenia Bulletin*, 17, 113-132.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed.) — *Text revision*. Washington, DC: Author.
- American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed.) — Text revision. Washington, DC: Author.
- Andelman, F., Fried, I., & Neufeld, M. Y. (2001). Quality of life self-assessment as a function of lateralization of lesion in candidates for epilepsy surgery. *Epilepsia*, 42, 549-555.
- Arbisi, P. A. & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale, *F(p). Psychological Assessment, 7*, 424-431.

- Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. (2002). A comparison of MMPI-2 validity in African American and Caucasian psychiatric inpatients. *Psychological Assessment*, 14, 3-15.
- Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2003). Empirical correlates of common MMPI-2 two-point codes in male psychiatric inpatients. *Assessment*, 10, 237-247.
- Archer, J. (1999). Assessment of the reliability of the conflict tactics scale: A metaanalytic review. *Journal of Interpersonal Violence, 14*, 1263-1289.
- Archer, R. P. (2009). *MMPI-A: Assessing adolescent psychopathology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Archer, R. P., Fontaine, J., & McCrae, R. R. (1998). Effects of two MMPI-2 validity scales on basic scale relations to external criteria. *Journal of Personality Assessment*, 70, 87-102.
- Arnold, H., Feldman, D., & Purbhoo, M. (1985). The role of social-desirability response bias in turnover research. *Academy of Management Journal*, 28, 955-966.
- Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The* Alcohol Use Disorders Identification Test: Guidelines for use in primary care – Second edition. World Health Organization – Department of Mental Health and Substance Dependence, Geneva.

- Bagby, R. M., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999).Defensive responding on the MMPI-2 in family custody and access evaluations.*Psychological Assessment, 11,* 24-28.
- Ballard, C. G., Chithiramohan, R. N., Handy, S., Bannister, C., Davis, R., & Todd, N. B.
 (1991). Information reliability in dementia sufferers. *International Journal of Geriatric Psychiatry*, 6, 313-316.
- Baer, R. A., Ballenger, J., & Kroll, L. S. (1998). Detection of underreporting on the MMPI-A in clinical and community samples. *Journal of Personality Assessment*, 71, 98-113.
- Baer, R. A. & Sekirnjak, G. (1997). Detection of underreporting on the MMPI-2 in a clinical population: Effects of information about validity scales. *Journal of Personality Assessment, 69*, 555-567.
- Bash, I. Y. & Alpert, M. (1980). The determination of malingering. Annals of the New York Academy of Sciences, 347, 86-99.
- Basler, H. D. (2007). Peculiarities of pain assessment and treatment in the elderly. Verhaltenstherapie & Verhaltensmedizin, 28, 398-406.
- Basso, M. R., Carona, F. D., Lowery, N., & Axelrod, B. N. (2002). Practice effects on the WAIS-III across 3- and 6-month intervals. *The Clinical Neuropsychologist*, 16, 57-63.
- Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI-2 in child custody litigation. *Psychological Assessment*, *9*, 205-211.

- Ben-Porath, Y. S. (2003). Assessing personality and psychopathology with self-report inventories. In I. B. Weiner (Series Ed.) & A. M. Goldstein (Vol. Ed.), *Handbook* of psychology: Vol. 11. Forensic psychology (pp. 485-508). Hoboken, NJ: Wiley.
- Ben-Porath, Y. S., Graham, J. R. & Tellegen A. (2009). The MMPI-2 Symptom Validity (FBS) Scale Development, Research Findings, and Interpretive Recommendations. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y. S. & Tellegen, A. (2008). MMPI-2-RF manual for administration, scoring, and interpretation. Minneapolis: University of Minnesota Press.
- Berry, D. T. R., Wetter, M. W., Baer, R. A., Youngjohn, J. R., Gass, C. S., Lamb, D. G., et al. (1995). Overreporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment*, 7, 517-523.
- Binder, L. M. & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentive on recovery after closed-head injury. *American Journal of Psychiatry*, 153, 7-10.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A.(2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317-335.
- Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI.* Norwalk, CT: Appleton-Century-Crofts.
- Bond, J. & Cherpitel, C. J. (2004). Measuring accuracy of self-reported vehicular offences and blood alcohol concentration levels among whites and Mexican

Americans mandated to drinking driver programs. *Addictive Behaviors*, *29*, 1163-1169.

- Bordini, E. J., Chaknis, M. M., Eckman-Turner, R. M., & Perna, R. B. (2002). Advances and issues in the diagnostic differential of malingering versus brain injury. *Neurorehabilitation*, 17, 93-104.
- Braff, D. L., Heaton, R., Kuck, J., Cullum, M., Moranville, J., Grant, I., et al. (1991). The generalized pattern of neuropsychological deficits in outpatients with chronic schizophrenia with heterogeneous Wisconsin Card Sorting Test results. *Archives* of General Psychiatry, 48, 891-898.
- Briere, J. (1996). Trauma Symptom Checklist for Children (TSCC) Professional Manual.Odessa, FL: Psychological Assessment Resources.
- Briere, J. & Elliott, D. M. (1997). Psychological assessment of interpersonal victimization effects in adults and children. *Psychotherapy*, *34*, 353-364.
- Bruck, M., Ceci, S. J., & Hembrooke, H. (1998). Reliability and credibility of young children's reports. *American Psychologist*, *53*, 136-151.
- Burchett, D. L. & Ben-Porath, Y. S. (2010). The impact of overreporting on MMPI-2-RF substantive scale score validity. *Assessment*, *17*, 497-516.
- Buss, A. H. & Plomin, R. (1975). A Temperament Theory of Personality Development. John Wiley & Sons: New York.

- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2 Manual for Administration, Scoring, and Interpretation (Rev. ed.)*. Minneapolis: University of Minnesota Press.
- Butler, R. N. (1978). The doctor and the aged patient. In W. Reichel (Ed.), *The geriatric patient*. New York: HP Publishing.
- Bylin, S. & Christianson, S. A. (2002). Characteristics of malingered amnesia:Consequences of withholding vs. distorting information on later memory of a crime event. *Legal and Criminological Psychology*, *13*, 495-511.
- Caetano, R. & Clark, C. L. (2000). Hispanics, Blacks, and Whites driving under the influence of alcohol: Results from the 1995 National Alcohol Survey. Accident Analysis and Prevention, 32, 57-64.
- Cashel, M. L., Rogers, R., Sewell, K., & Martin-Cannici, C. (1995). The Personality Assessment Inventory and the detection of defensiveness. *Assessment*, *2*, 333-342.
- Christianson, S. A. & Bylin, S. (1999). Does simulating amnesia mediate genuine forgetting for a crime event? *Applied Cognitive Psychology*, *13*, 495-511.
- Clark, C. R. (1988). Sociopathy, malingering and defensiveness. In R. Rogers (Ed.),
 Clinical assessment of malingering and deception (pp. 55-64). New York:
 Guilford.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, *5*, 115-124.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression / correlation analyses for the behavioral sciences – Third Edition. Mahwah, NJ: Erlbaum.
- Cohen, R. J. & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). New York: McGraw Hill.
- Coleman, R. D., Rapport, L. J., Millis, S. R., Ricker, J. H., & Farchione, T. J. (1998).
 Effects of coaching on detection of malingering on the California Verbal Learning
 Test. *Journal of Clinical and Experimental Neuropsychology*, 20, 201-210.
- Costa, P. T. & McCrae, R. R. (1985). Hypochondriasis, neuroticism, and aging. *American Psychologist*, 40, 19-28.
- Costa, P. T. & McCrae, R. R. (1992). *NEO-PI-R and NEO-FFI: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Craig, R. J. (2006). The Millon Clinical Multiaxial Inventory III. In R. P. Archer (Ed.),
 Forensic uses of clinical assessment instruments (pp. 121-146). Mahwah, NJ:
 Lawrence Erlbaum Associates.
- Crowne, D. P. & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Cunnien, A. J. (1988). Psychiatric and medical syndromes associated with deception. InR. Rogers (Ed.), *Clinical assessment of malingering and deception* (pp. 13-33).New York: Guilford.

- Czaja, R., Blair, J., Bickart, B., & Eastman, E. (1994). Respondent strategies for recall of crime victimization incidents. *Journal of Official Statistics*, 10, 257-276.
- Dalby, J. T. (1988). Detecting faking in the pretrial psychological assessment. *American Journal of Forensic Psychology*, *2*, 49-55.

DeBettignies, B. H., Mahunn, R. K., & Pirozzlo, F. J. (1990). Insight for impairment in independent living skills in Alzheimer's disease and multi-infarct dementia. *Journal of Clinical and Experimental Neuropsychology*, 12, 355-363.

- Derogatis, L. R. (1994). *SCL-90-R Administration, Scoring, and Procedural Manual Third Edition*. Minneapolis: Pearson Assessments.
- Derouesné, C., Dealberto, M. J., Boyer, P., Lubin, S., Sauron, B., Piette, F., et al. (1993).
 Empirical evaluation of the 'Cognitive Difficulties Scale' for assessment of memory complaints in general practice: A study of 1628 cognitively normal subjects aged 45-75 years. *International Journal of Geriatric Psychiatry*, *8*, 599-607.
- Dixon, M. J. & King, S. (1995). The concordance between symptom information gathered from remitted schizophrenic patients and their relatives. *Journal of Psychiatric Research*, 29, 447-456.
- Drob, S. L., Meehan, K. B., & Waxman, S. E. (2009). Clinical and conceptual problems in the attribution of malingering in forensic evaluations. *The Journal of the American Academy of Psychiatry and the Law, 37*, 98-106.

- Duberstein, P. R. & Heisel, M. J. (2007). Personality traits and the reporting of affective disorder symptoms in depressed patients. *Journal of Affective Disorders*, 103, 165-171.
- Eaton, W. W., Neufeld, K., Chen, L. S., & Cai, G. (2000). A comparison of self-report and clinical diagnostic interviews for depression: Diagnostic Interview Schedules for clinical assessment in neuropsychiatry in the Baltimore Epidemiological Catchment Area follow-up. *Archives of General Psychiatry*, *57*, 217-222.
- Edens, J. & Ruiz, M. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment*, 18, 220-224.
- Ehrman, R. N., Robbins, S. J., & Cornish, J. W. (1997). Comparing self-reported cocaine use with repeated urine tests in outpatient cocaine abusers. *Experimental and Clinical Psychopharmacology*, 5, 150-156.
- Elhai, J. D., Gold, P. B., Frueh, B. C., & Gold, S. N. (2000). Cross-validation of the MMPI-2 in detecting malingered posttraumatic stress disorder. *Journal of Personality Assessment*, 75, 449-463.
- Elhai, J. D., Gold, S. N., Sellers, A. H., & Dorfman, W. I. (2001). The detection of malingered posttraumatic stress disorder with MMPI-2 fake bad indices. *Assessment*, 8, 221-236.
- Elliott, D. M. & Briere, J. (1994). Forensic sexual abuse evaluations of older children: Disclosures and symptomatology. *Behavioral Sciences & the Law, 12,* 261-277.

- Eysenck, M. W. & Derakshan, N. (1999). Self-reported and other-rated trait anxiety and defensiveness in repressor, low-anxious, high-anxious, and defensive highanxious groups. *Anxiety, Stress & Coping: An International Journal, 12*, 127-144.
- Eysenck, H. J. & Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London: University Press.
- Fendrich, M., Mackesey-Amiti, M. E., Johnson, T. P., Hubbell, A., & Wislar, J. S.
 (2005). Tobacco-reporting validity in an epidemiological drug-use survey. *Addictive Behaviors*, 30, 175-181.
- Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, 10, 87-99.

Field, A. Discovering Statistics Using SPSS – Third Edition. London: SAGE.

- Frederick, R. I. & Bowden, S. C. (2008). The test validation summary. *Assessment, 16,* 181-192.
- Frederick, R. I. & Bowden, S. C. (2009). Evaluating constructs represented by symptom validity tests in forensic neuropsychological assessment of traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 24, 104-121.
- Fricker, J., Baelde, D., Igoin-Apfelbaum, L, Huet, J. M., & Apfelbaum, M. (1992).Underreporting of food intake in obese "small eaters." *Appetite*, *19*, 273-283.
- Frueh, B. C., Gold, P. B., & de Arellano, M. A. (1997). Symptom overreporting in combat veterans evaluated for PTSD: Differentiation on the basis of compensation seeking status. *Journal of Personality Assessment*, 68, 369-384.

- Frueh, B. C., Hamner, M. B., Cahill, S. P., Gold, P. B., & Hamlin, K. L. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD. *Clinical Psychology Review*, 20, 853-885.
- Frueh, B. C., Smith, D. W., & Barker, S. E. (1996). Compensation seeking status and psychometric assessment of combat veterans seeking treatment for PTSD. *Journal* of Traumatic Stress, 9, 427-439.
- Geraerts, E., Jelcic, M., & Merckelbach, H. (2006). Symptom overreporting and recovered memories of childhood sexual abuse. *Law & Human Behavior*, *30*, 621-630.
- Ghitza, U. E., Epstein, D. H., & Preston, K. L. (2007). Nonreporting of cannabis use: Predictors and relationship to treatment outcome in methadone maintained patients. *Addictive Behaviors*, *32*, 938-949.
- Gold, L. H. (2004). The workplace. In R. I. Simon & L. H. Gold (Eds.), *The American psychiatric textbook of forensic psychiatry* (pp. 303-326). Washington, DC:
 American Psychiatric Press.
- Gontkovsky, S. T. (1998). Huntington's disease: A neuropsychological overview. *The Journal of Cognitive Rehabilitation*, *16*, 6-9.
- Graham, J. R. (2006). *MMPI-2: Assessing Personality and Psychopathology Fourth Edition*. New York: Oxford.
- Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). MMPI-2 correlates for outpatient community mental health settings. Minneapolis, MN: University of Minnesota Press.
- Gray, T. A. & Wish, E. D. (1999). Correlates of underreporting recent drug use by female arrestees. *Journal of Drug Issues*, *29*, 91-105.
- Green, J., Goldstein, F. C., Sirockman, B. E., & Green, R. C. (1993). Variable awareness of deficits in Alzheimer's disease. *Neuropsychiatry*, *Neuropsychology*, and *Behavioral Neurology*, 6, 159-165.
- Greiffenstein, M. F. & Baker, W. J. (2001). Comparison of premorbid and postinjury MMPI-2 profiles in late preconcussion claimants. *The Clinical Neuropsychologist*, 15, 162-170.
- Griffin, B. Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. *Personality and Individual Differences*, *36*, 1545-1558.
- Gruzelier, J., Seymour, K., Wilson, L., Jolley, A., & Hirsch, S. (1988). Impairments on neuropsychological tests of temporohippocampal and frontohippocampal functions and word fluency in remitting schizophrenia and affective disorders. *Archives of General Psychiatry*, 45, 623-629.
- Hamby, S. L. (2005). Measuring gender differences in partner violence: Implications from research on other forms of violent and socially undesirable behavior. *Sex Roles, 52*, 725-742.

- Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment, 22,* 87-95.
- Hathaway, S. R. & McKinley, J. C. (1940). A multiphasic personality schedule(Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249-254.
- Hay, G. G. (1983). Feigned psychosis A review of the simulation of mental illness. British Journal of Psychiatry, 143, 8-10.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer Academic / Plenum Publishers.
- Heinze, M. C. (1999). "Yet there's method in his madness...": Dimensions of deception and dangerousness. *Aggression and Violent Behavior*, *4*, 387-412.
- Hilts, D. & Moore Jr., J. M. (2003). Normal range MMPI-A profiles among psychiatric inpatients. *Assessment*, 10, 266-272.
- Hirschi, T., Hindelang, M. J., & Weis, J. G. (1980). The status of self-report measures. InM. W. Klein and K. S. Teilman, eds., *Handbook of Criminal Justice Evaluation*.Beverly Hills, Sage Publications.
- Hodgins, D. C. & Makarchuk, K. (2003). Trusting problem gamblers: Reliability and validity of self-reported gambling behavior. *Psychology of Addictive Behaviors*, 17, 244-248.

- Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, 51, 469-477.
- Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior and Personality*, *13*, 387-398.
- Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal* of Behavioural Science, 39, 184-201.
- Holden, R. R. & Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment*, 4, 170-173.
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63, 272-279.
- Hong, S. & Roznowski, M. (2001). An investigation of the influence of test bias on regression slope. *Applied Measurement in Education*, 14, 351-368.
- Hoth, K. F., Paulsen, J. S., Moser, D. J., Tranel, D., Clark, L. A., & Bechara, A. (2007).
 Patients with Huntington's disease have impaired awareness of cognitive, emotional, and functional abilities. *Journal of Clinical and Experimental Neuropsychology*, 29, 365-376.

- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990).Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581-595.
- Hser, Y. I., Maglione, M., & Boyle, K. (1999). Validity of self-report of drug use among STD patients, ER patients, and arrestees. *American Journal of Drug and Alcohol Abuse*, 25, 81-91.
- Huizinga, D. & Elliott, D. S. (1986). Reassessing the reliability and validity of self-report delinquency measures. *Journal of Quantitative Criminology*, 2, 293-327.
- Hyer, L., Boudewyns, P., Harrison, W. R., O'Leary, W. C., Bruno, R., D., Saucer, R. T., et al. (1988). Vietnam veterans: Overreporting versus acceptable reporting of symptoms. *Journal of Personality Assessment*, 52, 475-486.
- Iverson, G. L. & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma and Rehabilitation*, 15, 829-858.
- Ja, D. Y. & Aoki, B. (1993). Substance abuse treatment: Cultural barriers in the Asian-American community. *Journal of Psychoactive Drugs*, 25, 61-71.
- Jaccard, J., & Turrisi, R. (2003). Interaction effects in multiple regression, 2nd edition. Thousand Oaks, CA: Sage.
- Jackson, D. N., Wrobleski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance*, 13, 371-388.

- James, W. H., Kim, G. K., & Moore, D. D. (1997). Examining racial and ethnic differences in Asian Adolescent drug use: The contributions of culture, background and lifestyle. *Drugs: Education, Prevention and Policy, 4*, 39-51.
- Ju, D. & Varney, N. R. (2000). Can head injury patients simulate malingering? Applied Neuropsychology, 7, 201-207.
- Kashdan, T. B., Elhai, J. D., & Frueh, B. C. (2007). Anhedonia, emotional numbing, and symptom overreporting in male veterans with PTSD. *Personality and Individual Differences*, 43, 725-735.
- Kim, M. T. & Hill, M. N. (2003). Validity of self-report of illicit drug use in young hypertensive urban African American males. *Addictive Behaviors*, 28, 795-802.
- Klesges, R. C., Eck, L. H., & Ray, J. W. (1995). Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition
 Examination Survey. *Journal of Consulting and Clinical Psychology*, 63, 438-444.
- Kopelman, M. D. (1995). The assessment of psychogenic amnesia. In A. D. Baddeley, B.A. Wilson, & F. N. Watts (Eds.), *Handbook of memory disorders* (pp. 427-448).New York: John Wiley.
- Kropp, P. R. & Rogers, R. (1993). Understanding malingering: Motivation, method, and deception. In M. Lewis & C. Saarni (Eds)., *Lying and deception in everyday life* (pp. 201-216). New York: Guilford Press.

- Kurtz, J. E. & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, 76, 315-332.
- Kurtz, J. E., Tarquini, S. J., & Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences*, 45, 22-27.
- Lamb, D. G., Berry, D. T. R., Wetter, M. W., & Baer, R. A. (1994). Effects of two types of information on malingering of closed head injury on the MMPI-2: An analog investigation. *Psychological Assessment*, 6, 8-13.
- Lanyon, R. I. & Almer, E. R. (2002). Characteristics of compensable disability patients who choose to litigate. *Journal of the American Academy of Psychiatry and the Law, 30*, 400-404.
- Lanyon, R. I. & Cunningham, K. S. (2005). Construct validity of the misrepresentation scales of the Psychological Screening Inventory. *Journal of Personality Assessment*, 85, 197-206.
- Lapham, S. C., C'de Baca, J., McMillan, G., & Hunt, W. C. (2004). Accuracy of alcohol diagnosis among DWI offenders referred for screening. *Drug and Alcohol Dependence*, 76, 135-141.
- Lara, J. J., Scott, J. A., & Lean, M. E. J. (2004). Intentional mis-reporting of food consumption and its relationship with body mass index and psychological scores in women. *Journal of Human Nutrition and Dietetics*, 17, 209-218.

- Ledgerwood, D. M., Goldberger, B. A., Risk, N. K., Lewis, C. E., & Price, R. K. (2008).Comparison between self-report and hair analysis of illicit drug use in a community sample of middle-aged men. *Addictive Behaviors, 33*, 1131-1139.
- Lees-Haley, P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs:Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, *53*, 745-755.
- 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*, 68, 203-201.
- Li, A. & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131-141.
- Lilienfeld, S.O. & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, 66, 488–524.
- Lu, N. T., Taylor, B. G., & Riley, K. J. (2001). The validity of adult arrestee self-reports of crack cocaine use. *American Journal of Drug and Alcohol Abuse*, *27*, 399-419.
- Lukoff, D., Nuecheterlein, K. H., & Venura, J. (1986). Manual for the expanded brief psychiatric rating scale. *Schizophrenia Bulletin*, *12*, 578-602.
- Lynam, D. R., Whiteside, S., & Jones, S. (1999). Self-reported psychopathy: A validation study. *Journal of Personality Assessment*, 73, 110-132.

- Mahar, D., Colognon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, 18, 605-609.
- Martin, P. J., Hunter, M. L., & Moore, J. E. (1977). Pulling the wool: Impressionmanagement among hospitalized schizophrenics. *Research Communications in Psychology, Psychiatry, and Behavior, 2*, 21-26.
- McCrae, R. R. & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*, 882-888.
- McCrae, R. R., Stone, S. V., Fagan, P. J., & Costa, P. T. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality*, 66, 285-313.
- 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. C., Hathaway, S. R., & Meehl, P. E. (1948). The Minnesota Multiphasic
 Personality Inventory: VI. The K scale. *Journal of Consulting Psychology*, 12, 20-31.
- McNair, D. M. & Kahn, R. J. (1983). Self-assessment of cognitive deficits. In Assessment in Geriatric Psychopharmacology (T. Crook, S. Ferris and R. Bartus, Eds.). Mark Powley Associates: New Cannan, CT.

- Meehl, P. E. & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. *Journal of Applied Psychology*, *30*, 525-564.
- Mercado, M. M. (2000). The invisible family: Counseling Asian American substance abusers and their families. *The Family Journal: Counseling and Therapy for Couples and Families*, 8, 267-272.
- Mersman, J. L. & Schulz, K. S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences*, *24*, 217-227.
- Messina, N. P., Wish, E. D., Nemes, S., & Wraight, B. (2000). Correlates of underreporting of post-discharge cocaine use among therapeutic community clients. *Journal of Drug Issues*, 30, 119-132.
- Midanik, L. (1982). Over-reports of recent alcohol consumption in a clinical population:A validity study. *Drug and Alcohol Dependence*, 9, 101-110.
- Miller, A. R. & Rosenfeld, J. P. (2003). Response-specific scalp distributions in deception detection and ERP correlates of psychopathic personality traits. *Journal* of Psychophysiology, 18, 13-26.
- Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test (M-FAST): Professional manual.* Odessa FL: Psychological Assessment Resources.
- Millon, T. (1983). *Millon Clinical Multiaxial Inventory*. Minneapolis: National Computer Systems.
- Millon, T. (1987). *Millon Clinical Multiaxial Inventory—II manual*. Minneapolis, MN: National Computer Systems.

- Millon, T. (1994). *Millon Clinical Multiaxial Inventory—III manual*. Minneapolis, MN: National Computer Systems.
- Millon, T., Davis, R., & Millon, C. (1997) Millon Clinical Multiaxial Inventory—III manual (2nd ed.). Minneapolis, MN: National Computer Systems.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102.
- Morel, K. R. (1998). Development and preliminary validation of a forced-choice test of response bias for post-traumatic stress disorder. *Journal of Personality Assessment*, 67, 244-257.
- Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey (1993, August). *Defensiveness and malingering indices for the PAI*. Paper presented at the annual convention of the American Psychological Association, Toronto, Ontario, Canada.
- Morey, L. C. (1996). An interpretive guide to the Personality Assessment Inventory. Odessa, FL: Personality Assessment Resources.
- Morrall, A. R., McCafrey, D., & Iguchi, M. Y. (2000). Hardcore drug users claim to be occasional users: Drug use frequency underreporting. *Drug and Alcohol Dependence*, 57, 193-202.

- O'Connor, M. G. & Lafleche, G. (2006). Amnestic syndromes. In P. J. Snyder, P. D.
 Nussbaum, & D. L. Robins (Eds.), *Clinical neuropsychology: A pocket handbook for assessment* (pp. 463-488). Washington, DC: American Psychological Association.
- Ones, D. S. & Viswesveran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245-269.

Pampel, F. C. (2000). Logistic regression. Thousand Oaks, CA: Sage.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal* of Personality Assessment, 46, 598-609.
- Paulhus, D. L. (1988). Assessing self-deception and impression management in selfreports: The Balanced Inventory of Desirable Responding. Unpublished manual, University of British Columbia, Vancouver, Canada.
- Pauls, C. A. & Crost, N. W. (2005). Cognitive ability and self-reported efficacy of selfpresentation predict faking on personality measures. *Journal of Individual Differences*, 26, 194-206.
- Petroskey, L. J., Ben-Porath, Y. S., & Stafford, K. P. (2003). Correlates of the Minnesota Multiphasic Personality Inventory—2 (MMPI-2) Personality Psychopathology
 Five (PSY-5) scales in a forensic assessment setting. *Assessment, 10*, 393-399.

- Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology*, 78, 582-593.
- Pollock, P. (1996). A cautionary note on the determining of malingering in offenders. *Psychology, Crime and Law, 3*, 97-110.
- Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators on the Test of Memory Malingering. *Archives of General Neuropsychology*, 19, 693-702.
- Reid-Seiser, H. L., & Fritzche, B. A. (2001). The usefulness of the NEO PI-R Positive Presentation Management scale for detecting response distortion in employment contexts. *Personality and Individual Differences*, *31*, 639-650.
- Rees, C. J. & Metcalfe, N. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology*, 18, 156-165.
- Robie, C., Curtin, P. J., Foster, C., Philips, H. L., Zbylut, M., & Tetrick, L. E. (2000). The effect of coaching on the utility of response latencies in detecting fakers on a personality measure. *Canadian Journal of Behavioural Science*, 32, 226-233.
- Rockett, I. R. H., Putnam, S. L., Jia, H., & Smith, G. S. (2006). Declared and undeclared substance use among emergency department patients: A population-based study. *Addiction*, 101, 706-712.
- Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21, 182-188.

- Rogers, R. (1992). Structured Interview of Reported Symptoms. Odessa, FL: Psychological Assessment Resources.
- Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67, 629-640.
- Rogers, R. (2008a). An introduction to response styles. In R. Rogers (Ed.), *Clinical* assessment of malingering and deception (3rd ed.) (pp. 3-13). New York: Guilford.
- Rogers, R. (2008b). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.) (pp. 14-38).
 New York: Guilford.
- Rogers, R. (2008c). Researching response styles. In R. Rogers (Ed.), *Clinical assessment* of malingering and deception (3rd ed.) (pp. 411-434). New York: Guilford.
- Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Feigning schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment*, 60, 215-226.
- Rogers, R. & Bender, S. D. (2003). Evaluation of malingering and deception. In I. B.
 Weiner (Series Ed.) & A. M. Goldstein (Vol. Ed.), *Handbook of psychology: Vol.*11. Forensic psychology (pp. 109-132). Hoboken, NJ: Wiley.
- Rogers, R. & Cavanaugh, J. L. (1983). "Nothing but the truth"...A reexamination of malingering. *Journal of Law and Psychiatry*, 11, 443-460.

- Rogers, R. & Dickey, R. (1991). Denial and minimization among sex offenders: A review of competing models of deception. *Annals of Sex Research*, *4*, 49-63.
- Rogers, R., Gillis, J. R., & Bagby, R. M. (1990). The SIRS as a measure of malingering:
 A validation study with a correctional sample. *Behavioral Sciences & the Law*, 8, 85-92.
- Rogers, R., Gillis, J. R., Bagby, R. M., & Montiero, E. (1991). Detection of malingering on the Structured Interview of Reported Symptoms (SIRS): A study of coached and uncoached simulators. *Psychological Assessment*, *3*, 673-677.
- Rogers, R., Kropp, P. R., Bagby, R. M., & Dickens, S. E. (1992). Faking specific disorders: A study of the Structured Interview of Reported Symptoms (SIRS). *Journal of Clinical Psychology*, 48, 643-648.
- Rogers, R., Ornduff, S. R., & Sewell, K. W. (1993). Feigning specific disorders: A study of the Personality Assessment Inventory (PAI). *Journal of Personality Assessment, 60, 554-560.*
- Rogers, R. & Payne, J. W. (2006). Damages and rewards: Assessment of malingered disorders in compensation cases. *Behavioral Sciences and the Law*, 24, 645-658.
- Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (1998). A comparison of forensic and non-forensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior*, 22, 353-367.
- Rogers, R., Sewell, K. W., & Goldstein, A. M. (1994). Explanatory models of1malingering: A prototypical analysis. *Law and Human Behavior*, 18, 543-552.

- Rogers, S., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment*, 67, 629-640.
- Romano, E. & De Luca, R. V. (2001). Male sexual abuse: A review of effects, abuse characteristics, and links with later psychological functioning. *Aggression and Violent Behavior*, 6, 55-78.
- Rosay, A. B., Najaka, S. S., & Herz, D. C. (2007). Differences in the validity of selfreported drug use across five factors: Gender, race, age, type of drug, and offense seriousness. *Journal of Quantitative Criminology*, 23, 41-58.
- Rosenberg, M. (1989). Society and the Adolescent Self-Image. Wesleyan University Press: Middletown, CT.
- Scholz, U., Doña, B. C., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal* of Psychological Assessment, 18, 242-251.
- Schwarzer, R. (2008). Website PDF: Everything you wanted to know about the General Self-Efficacy Scale but were afraid to ask. July 7, 2008. Retrieved August 7, 2008.
- Schwarzer, R. & Born, A. (1997). Optimistic self-beliefs: Assessment of general perceived self-efficacy in thirteen cultures. *World Psychology*, *3*, 177-190.

- Sellbom, M. (2007). Locating Psychopathy in Abnormal Range Personality Dimensions: The MMPI-2 Restructured Clinical (RC) Scales as a Road Map. Unpublished doctoral dissertation. Kent State University, Kent, OH.
- Sellbom, M. & Bagby, R. M. (2008). Response styles on multiscale inventories. In R.
 Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.) (pp. 182-206). New York: Guilford.
- Sellbom, M., & Verona, E. (2004, October). *Differential Neuropsychological Mechanisms for Two Psychopathy Factors*. Poster presented at the 19th Annual
 Meeting for the Society for Research in Psychopathology, St. Louis, MO.
- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEOPI-R: Development and initial validation. *Journal of Personality Assessment*, 68, 127-138.
- Shum, D. H. K., O'Gorman, J. G., & Alpar, A. (2004). Effects of incentive and preparation time on performance and classification accuracy of standard and malingering-specific memory tests. *Archives of General Neuropsychology*, 19, 817-823.
- Sinclair, P. A., Lyness, J. M., King, D. A., Cox, C., & Caine, E. D. (2001). Depression and self-reported functional status in older primary care patients. *American Journal of Psychiatry*, 158, 416-419.

Skinner, H. (1982). The Drug Abuse Screening Test. Addictive Behaviors, 7, 363-371.

- Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.) (pp. 323-342). New York: Guilford.
- Smith, G. P. & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the Academy of Psychiatry and Law*, 25, 183-189.
- Smyth, K. A., Neundorfer, M. M., Koss, E., Geldmacher, D. S., Ogrocki, P. K., & Whitehouse, P. J. (2002). Quality of life and deficit identification in dementia. *Dementia*, 1, 345-358.
- Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, 9, 219-242.
- Sommers, M. S., Dyehouse, J. M., Howe, S. R., Wekselman, K., & Fleming, M. (2002). "Nurse, I only had a couple of beers": Validity of self-reported drinking before serious vehicular injury. *American Journal of Critical Care*, 11, 106-114.
- Stein, L. A. R., Graham, J. R., & Williams, C. L. (1995). Detecting fake-bad MMPI-A profiles. *Journal of Personality Assessment*, 65, 415-427.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family*, 41, 75-88.

- Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child custody litigants. *Journal of Personality Assessment*, 73, 1-18.
- Stutts, J. T., Hickey, S. E., & Kasdan, M. L. (2003). Malingering by proxy: A form of pediatric condition falsification. *Developmental and Behavioral Pediatrics*, 24, 276-278.
- Suhr, J. A. & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), Assessment of malingered neuropsychological deficits (pp. 287-311). New York: Oxford.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics, 5th ed.* Boston: Allyn and Bacon.
- Tellegen, A. (1978). *Brief Manual for the Multidimensional Personality Questionnaire*. Unpublished manuscript. Minneapolis, MN: University of Minnesota.
- Tellegen, A. (1982). Manual for the Multidimensional Personality Questionnaire.Unpublished manuscript, Minneapolis, MN: University of Minnesota.
- Tett, R. P., Anderson, M. G., Ho, C.-L., Yang, T. S., Huang, L., & Hanvongse, A. (2006).
 Seven nested questions about faking in personality tests. In Griffith & Peterson
 (Eds.) *A closer examination of applicant faking behavior*. Greenwich, CT:
 Information Age Publishing.
- Tolin, D. F., Maltby, N., Weathers, F. W., Litz, B. T., Knight, J., & Keane, T. M. (2004). The use of the MMPI-2 Infrequency-Psychopathology Scale in the assessment of

posttraumatic stress disorder in male veterans. *Journal of Psychopathology and Behavioral Assessment, 26,* 23-29.

- Turner, A. G. (1972). The San Jose methods test of known crime victims. Washington, DC: U.S. Government Printing Office.
- van Oorsouw, K. & Merckelbach, H. (2004). Feigning amnesia undermines memory for a mock crime. *Applied Cognitive Psychology*, *18*, 505-518.
- van Oorsouw, K. & Merckelbach, H. (2006). Simulating amnesia and memories of a mock crime. *Psychology, Crime, & Law, 12,* 261-271.
- Vansant, G. & Hulens, M. (2006). The assessment of dietary habits in obese women: Influence of eating behavior patterns. *Eating Disorders, 14*, 121-129.
- Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., et al. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology*, 19, 37-48.
- Walters, G. D. (2006). Coping with malingering and exaggeration of psychiatric symptomatology in offender populations. *American Journal of Forensic Psychology*, 24, 21-40.
- Weinborn, M., Orr, T., Woods, S. P., Conover, E., & Feix, J. (2003). A validation of the Test of Memory Malingering in a forensic psychiatric setting. *Journal of Clinical* and Experimental Neuropsychology, 25, 979-990.

- Widows, M. & Smith, G. P. (2005). Structured Inventory of Malingered Symptomatology (SIMS) and professional manual. Odessa, FL: Psychological Assessment Resources.
- Yates, D., Moore, D., & McCabe, G. (1999). *The practice of statistics (1st Ed.)*. New York: W.H. Freeman.
- Yudko, E., Lozhkina, O., & Fouts, A. (2007). A comprehensive review of the psychometric properties of the Drug Abuse Screening Test. *Journal of Substance Abuse Treatment*, 32, 189-198.