

EFFECTS OF DEFENSIVENESS ON THE REPORTING OF
PERSONALITY DISORDER SYMPTOMS

Chelsea E. Fiduccia, B.A.

Thesis Prepared for the Degree of
MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2011

APPROVED:

Richard Rogers, Major Professor
Adriel Boals, Committee Member
Jennifer L. Callahan, Committee Member
and Director of Clinical Training
Vicki L. Campbell, Chair of the Department
of Psychology
James D. Meernik, Acting Dean of the
Toulouse Graduate School

Fiduccia, Chelsea E. Effects of Defensiveness on the Reporting of Personality Disorder Symptoms. Master of Science (Psychology), December 2011, 121 pp., 22 tables, 2 figures, references, 130 titles.

Personality disorders are not granted the same clinical attention accorded Axis I disorders despite their instrumental role in treatment and outcome. Even when standardized assessments are used, their clinical utility may be limited by an overly favorable self-presentation. The current study focused on defensiveness, the intentional denial of symptomatology, by examining individuals' ability to minimize their presentation on personality disorder diagnostic measures. Using a within-subjects simulation design, dually diagnosed inpatients were assessed under both honest and defensive conditions. The study used self-report (Structured Clinical Interview for DSM-IV – Axis II – Personality Questionnaire, SCID-II-PQ) and interview-based (Structured Interview for DSM-IV Personality, SIDP-IV) diagnostic measures and a self-report measure of favorable self-presentation (Paulhus Deception Scales, PDS). The inpatients were quite capable of hiding maladaptive personality traits on diagnostic measures, with similarly large effect sizes on both the SCID-II-PQ and SIDP-IV. In addition to the PDS, two new detection strategies for identifying defensiveness showed promise.

Copyright 2011
by
Chelsea E. Fiduccia

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	viii
Chapters	
1. INTRODUCTION	1
Stigma	2
Lack of Insight	7
Models of Favorable Self-Presentation.....	8
Paulhus' Model	9
Lanyon's Model	11
Rogers' Model	13
Personality Disorder Assessment and Favorable Self-Presentation	14
Prevalence	16
Validity Scales	17
Defensiveness on the MCMI	21
Current Study	26
Research Questions and Hypotheses	27
Research Question 1	27
Research Question 2	27
Research Question 3	28
Research Question 4	28
Supplementary Question 1	29
Supplementary Question 2.....	29
2. METHODS	30

Design	30
Participants.....	31
Research and Administrative Approval	31
Measures	31
Structured Interview for DSM-IV Personality (SIDP-IV).....	31
Structured Clinical Interview for DSM-IV, Axis II, Personality Questionnaire (SCID-II-PQ)	32
Paulhus Deception Scales (PDS)	33
Procedures	35
Participant Recruitment	35
Informed Consent.....	35
Phase I.....	36
Phase II.....	36
3. RESULTS	38
Scoring and Reliability of Research Measures	38
PDS Scoring and Reliability	38
Reliability of Axis II Measures.....	40
Refinement of the Sample.....	42
Manipulation Check.....	42
Exclusion Criteria	42
Representativeness of Final Sample	43
Parametric Assumptions	45
Missing Data and Scoring Procedures	45
Axis II Symptomatology.....	46
Defensiveness on Personality Disorder Measures	48

Defensiveness on Paulhus Deception Scales	56
Cut Scores for Defensive Responding	57
Paulhus Deception Scales	58
Axis II Measures	60
Indiscriminate Denial of Symptoms (IDS)	61
Denial of Commonly Reported Symptoms (D-CRS)	63
Narcissism and Self-Deceptive Enhancement	65
Self-Reported Ability to Portray a Favorable Impression	65
4. DISCUSSION	67
The Role of Transparency and Structure in Personality Assessment	70
Favorable Self-Presentation in Personality Assessment	75
Defensive Responding on Personality Disorder Assessments	79
Susceptibility of Axis II Measures to Defensive Responding	79
Detection of Defensive Responding	83
Detection Strategies and Intentionality	85
Detection Strategies for Axis II Defensiveness	86
Limitations	87
Implications and Future Directions	89
APPENDICES	92
REFERENCES	105

LIST OF TABLES

1. Models of Favorable Self-Presentation.....10

2. Defensive Responding Effect Sizes on MCMI-II Validity and Axis II Diagnostic Scales25

3. Internal Consistency of PDS.....39

4. Internal Consistency of Personality Disorder Clusters and Diagnoses for SCID-II-PQ and
SIDP-IV41

5. Differences between Final Sample and Excluded Participants on Gender, Ethnicity, Marital
Status, and SES43

6. Differences between Final Sample and Excluded Participants on Age, GAF, and Highest
Grade Achieved44

7. Comorbidity of Personality Disorders among the Final Sample47

8. Axis II Symptomatology among the Final Sample of Dually Diagnosed Inpatients.....48

9. Effects of Condition, Measure, and Cluster on Reported Symptomatology49

10. Differences between Honest and Defensive Groups on Axis II Clusters and Diagnoses.....50

11. Axis II Symptomatology Reported under Defensive Instructions53

12. Differences between Honest and Defensive Conditions in Prevalence of Personality
Disorders55

13. Effects of Condition and Scale on PDS Scores57

14. Differences between Honest and Defensive Groups on PDS Scales57

15. Differences between Honest and Defensive Groups on PDS Scales59

16. Utility of IM Faking Good Cut Scores Classifying Defensiveness among Dually Diagnosed
Inpatients.....60

17. Utility of IDS Strategy Classifying Defensiveness among Dually Diagnosed Inpatients62

18. Commonly Reported Personality Disorder Symptoms Among Dually Diagnosed
 Inpatients.....64

19. Utility of SIDP-IV D-CRS Scale Classifying Defensiveness among Dually Diagnosed
 Inpatients.....64

20. Methods for Detecting Defensiveness on Axis II Diagnostic Measures77

21. Effects of Defensive Responding on Four Axis II Measures and Their Validity Scales.....81

22. Utility of Validity Scales on Four Axis II Measures in Clinical Samples87

LIST OF FIGURES

1. Proportion of reported Axis II symptomatology by condition, measure, and cluster51
2. Two-axis model of favorable self-presentation84

CHAPTER 1

INTRODUCTION

Since their official introduction into diagnostic taxonomy nearly 30 years ago, personality disorders have proved challenging to assess (Westen, 1997). As defined by the latest edition of the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; APA, 2000)*, these challenges include (a) their high rates of diagnostic comorbidity with other personality disorders and other mental disorders, and (b) great within-diagnosis heterogeneity (Verheul, 2005; Widiger, 1997, 2007; Zimmerman & Mattia, 1999). Despite challenges in diagnosis, the importance of maladaptive personality traits is undisputed and their potential severity warrants placement on a separate diagnostic axis (APA, 2000).

Axis II disorders are considered distinct from Axis I in terms of etiology, course, and treatment (APA, 2000). Mental retardation, also coded on Axis II, is clearly a lifelong developmental disorder that lacks effective treatments. In contrast to Axis II, Axis I disorders are viewed as generally episodic with more potential for successful treatment and remission, if not total amelioration, of symptoms. In this context, the view of personality disorders as lifelong "handicaps" is easily understood. However, evidence (Borge et al., 2010; Leichsenring & Leibing, 2003; Matusiewicz, Hopwood, Banducci, & Lejuez, 2010; Triebwasser & Siever, 2007) in recent years has suggested that personality disorders are not "life sentences," and symptoms can be successfully treated with both psychotherapy and pharmacotherapy.

Personality disorders, beyond perceptions of their chronicity and poor treatment outcomes, are often viewed negatively by both professionals and the public. According to Ganellen (2007), these maladaptive traits are often negatively viewed by others, leading some individuals to hide or minimize these characteristics; minimization may also be fueled by

internal distress relating to the symptoms. The disruption in interpersonal functioning, a primary feature of personality disorders, predisposes people with personality disorders to have difficult interactions with others.

Negative interpersonal interactions often give rise to labeling and judging based on social constructs, thus giving rise to stigmatization (Corrigan, Markowitz, & Watson, 2004; Link & Phelan, 2001; Overton & Medrina, 2008). Imposed by mental health professionals, pejorative labels (e.g., “narcissist” or “borderline”) have a significant harmful effect on individuals both internally and externally. Internally, labels can significantly damage individual’s self-efficacy and confidence (Overton & Medrina, 2008). Externally, labels also color the public’s perception, sometimes leading others to act in a hurtful manner toward those identified.

Stigma

Much research (Aviram, Brodsky, & Stanley, 2006; Link & Phelan, 2001, 2006) in the context of general mental disorders has focused on stigma, a constellation of negative attitudes toward an identifiable group of people that results in their broad debasement. Regarding psychological issues, Corrigan (2004) found mixed evidence regarding the specificity of stigma, with some researchers suggesting that any mental disorder is more stigmatized than general health and others suggesting that certain mental disorders are more highly stigmatized. In the latter view, disorders that evidence the strongest demarcation from normal behavior (e.g., psychotic disorders) are most stigmatized. Though personality disorders are less florid than psychotic disorders, individuals with personality disorders do manifest significant interpersonal disruptions and behavioral deviations far from the norm. However, stigmatization related to personality disorders has received very little research attention. Because of their off-putting interpersonal nature, clients with personality disorders are susceptible to stigmatizing attitudes.

To understand stigmatization, Corrigan (2004) identified two distinct types of stigma: public stigma and self-stigma. He and others (e.g., Overton & Medrina, 2008) define public stigma as the stereotypic beliefs of the general public regarding an individual's group membership (e.g., mental disorder), marked by an external evaluation based on societal norms, whereas self-stigma reflects a person's own negative beliefs about their group membership. Structural stigma is closely related to public stigma, but it is better conceptualized as the product of institutions rather than individuals (Corrigan et al., 2004; Link & Phelan, 2006). Corrigan and colleagues distinguished two types of structural discrimination: intentional and unintentional.

As a compelling example of intentional structural stigma, legislators in the United Kingdom moved, in 1999, for compulsory detainment of persons characterized with "dangerous severe personality disorder," regardless of whether they had been convicted of a crime (Beck, 2010; Kendell, 2002). To the informed mental health professional, this proposed legislation expressly contradicted England's 1998 Human Rights Act. That legislation explicitly prohibited detainment of unconvicted persons unless they met one of the following criteria: unsound mind, alcoholic, drug addict, vagrant, or likely to spread infectious diseases (Kendell, 2002). It appears, then, that public officials were equating having a personality disorder with being of unsound mind, spurring debate about whether personality disorders are severe mental disorders which warrant involuntary detainment and treatment. In describing the controversy surrounding this issue, Beck (2010) detailed how professionals were less in favor of indeterminate compulsory detainment than the public; however, this lack of support from professionals was not enough to stop the movement. This legislation, a reaction to a highly publicized violent crime, was ultimately dropped in favor of revising an existing act, though the outcome was nearly identical. Beck (2010) noted the emphasis on incapacitation rather than treatment, as male inmates with

little time left in their prison sentences are being transferred to specialized units for indeterminate commitment.

Stigma, whether based externally or internally, leads to shame, embarrassment, and diminished self-esteem (Corrigan, 2004; Link & Phelan, 2001, 2006; Overton & Medrina, 2008). Self-stigma is likely based on societal norms as well, but the difference occurs when negative evaluations and the associated feelings are internalized. As clinicians, it can be difficult to determine whether low self-efficacy and shame are stable characteristics or whether the client is experiencing the effects of self-stigma. As an approach most memorably heralded by Carl Rogers (1957), mental health professionals aim to provide for the client an accepting environment that is free from judgment, but this empathic approach may be difficult to achieve with clients who have personality disorders.

The interpersonal difficulties of people with personality disorders militate against forging effective treatment relationships. Though it might be subsumed as a facet of public stigma, a third type of stigma—generated by service providers—must be considered in clinical settings. Sartorius (2002) referred to this third type as “iatrogenic stigma” and noted that it further perpetuates both public and self-stigma. Only recently has this been a focus of research (see Schulze, 2007, for a review), though studies of iatrogenic stigma have tended to neglect personality disorders. Regarding iatrogenic stigma, treatment providers aspire to be unbiased and provide similar quality care to all patients that is uninfluenced by professionals’ negative attitudes. However, Aviram et al. (2006) found negative attitudes among health care workers toward patients with specific Axis II diagnoses, particularly Borderline Personality Disorder. These investigators suggested that the mechanism behind this stigmatization lies in the clinicians’ misattributions of the troublesome behaviors to these persons rather than to their

disorders; as a corollary, these professionals persistently believe that these persons can control their troublesome behaviors. Therefore, clinicians are likely to ascribe a willfulness to patients' behaviors that are, in theory, the direct result of their psychopathology.

Such negative attitudes can result in service providers withdrawing and providing less attention to the people with strongly stigmatized disorders (Aviram et al., 2006). In an informal survey of staff at a state psychiatric hospital, Sriram and Jabbarpour (2005) asked mental health staff members questions regarding their perceptions of schizophrenia. When faced with the hypothetical dilemma of disclosing their own diagnosis of schizophrenia, 60% of mental health staff acknowledged their unwillingness to disclose this diagnosis to nonprofessionals. Tellingly, 57% of those respondents indicated public stigma as their primary reason for nondisclosure. As evidenced by their study, mental health professionals are not immune to the stigmatization of mental disorders.

In introducing a special issue of the *Journal of Personality Disorders*, Clark (2006) argued that mental health professionals have difficulty separating their moral judgments about personality disorder criteria from their clinical perspective. With the conflation of personal and professional opinions, clinicians are likely to experience difficulty dealing with persons with Axis II diagnoses. However, relatively little research has explored clinicians' attitudes toward clients with personality disorders within the frame of iatrogenic stigma. In a seminal study on clinician stigma toward personality disorders, Lewis and Appleby (1988) examined psychiatrists' attitudes toward clients based on six case vignettes that differed on diagnosis, gender, or class. The psychiatrists were provided with a client vignette and asked to rate their attitudes on 22 client behaviors. More positive or accepting attitudes were rated lower than negative or judgmental statements. For example, an item related to overdose was anchored at "1" with the

statement “overdose would be a genuine suicidal act” and “6” with “overdose would be an attention-seeking act.” Clients described in the vignettes varied in gender, social class, and presence or absence of an unspecified personality disorder. Even when controlling for gender and social class, psychiatrists rated clients less favorably if their vignette even mentioned an Axis II diagnosis. In addition to attributing negative interpersonal traits, the psychiatrists generally expressed pessimistic views about treatment adherence and outcome for the personality disordered clients. Lewis and Appleby asserted that clinicians viewed these clients as difficult and unpleasant, based solely on the presence of a personality disorder.

Stigma related to personality disorders is perpetuated by self-fulfilling prophecies (Overton & Medrina, 2008). For example, if clinicians distance themselves emotionally or physically from a client with Borderline Personality Disorder, it might inadvertently serve to activate the client’s fear of rejection and result in additional acting-out behaviors. In this case, clinicians’ a priori beliefs about patients with personality disorders influence how the clinicians act and, thus, how the patients react. To avoid imposing clinicians’ expectations, Aviram et al. (2006) suggested that therapists use personal supervision to work on their biases and unduly negative expectations.

In an attempt to curtail iatrogenic stigma, Zimmerman, Rothschild, and Chelminski (2005) observed that some clinicians choose not to disclose Axis II diagnoses to their clients. This course of action seemingly protects clients from the public and self-stigma attached to the diagnosis. In a small British survey of mental health outpatients with personality disorders ($n = 10$) and their mental health service providers ($n = 9$), several clients had little or no idea about the concept of a “personality disorder” and also felt they had never received a clear explanation of their diagnoses (Stalker, Ferguson, & Barclay, 2005). Among patients aware of their

diagnoses, Corrigan (2004) contended that the different types of stigma could motivate a person to try to conceal their group membership. In the case of personality disorders, the concealment of these negatively viewed personality characteristics likely creates a major barrier to effective treatment. Clients cannot be expected to have insight into their personality problems if this diagnostic information is withheld from them.

Lack of Insight

Lack of insight is often described as one of the cardinal features of personality disorders. Most insight-related research has focused on psychotic disorders, but the conceptual definition of insight is applicable to any mental disorder, including Axis II diagnoses. Insight is separated into three related domains: awareness of the mental disorder, perceived need for treatment, and awareness of the consequences of the mental disorder (Amador et al., 1993). Because of a lack of awareness or insight into characteristics, Ganellen (2007) suggested that persons with Axis II traits might not be able to accurately describe their personality. Some individuals with personality disorders are genuinely unaware that they possess unfavorable personality characteristics (i.e., awareness of the mental disorder). Regardless of their awareness of the personality disorder, many individuals have difficulty identifying themselves as the source of their interpersonal problems and distress (i.e., awareness of the consequences of the mental disorder). Therefore, insight into personality disorders requires all three domains of insight: an awareness of the maladaptive traits, recognition of the consequences of those traits, and awareness of the need for treatment.

Much of the insight literature on psychotic disorders posits a causal relationship between lack of insight, poor treatment adherence, and poor outcomes. Lincoln, Lüllmann, and Rief's (2007) meta-analysis supports the first part of this causal chain that poor insight is associated

with poor treatment adherence in the short-term. Unexpectedly, they did not find a direct effect for insight on long-term adherence or functional outcomes. These causal links remain untested with personality disorders, but similar results are expected to predict short-term adherence. While clients with personality disorders often enter treatment for other reasons (e.g., Axis I disorders and acute stressors), many clinicians believe that insight into personality functioning is crucial to successful therapy (Bakken, Landheim, & Vaglum, 2007; Cacciola, Alterman, McKay, & Rutherford, 2001; Hayward, Slade, & Moran, 2006; Singer, 2005). The question of how insight might affect long-term functional outcomes, especially with the chronic nature of personality disorders (APA, 2000), has yet to be empirically examined.

As a result of stigmatization (public, self, or iatrogenic) and a concomitant lack of insight, the act of concealing negative personality characteristics can be influenced by situational as well as individual determinants. Lack of insight into maladaptive personality characteristics might lead to an inflexible, trait-like style of deceptive responding across settings. Public, self, and iatrogenic stigma, on the other hand, constitute situational demands that may elicit individualized deceptive responses based on specific circumstances. Several researchers and clinicians have attempted to delineate this group of response styles.

Models of Favorable Self-Presentation

The term “favorable self-presentation” is used to encompass responses that intentionally or unintentionally provide a substantially more positive impression of an individual than is warranted (Lanyon, 2004). Lanyon (2001) identified several other terms used to describe this positive style of responding; they include “simulated adjustment,” “faking good,” “socially desirable responding,” and “self-serving misrepresentation.” As noted in the varying terminology, there is no consensus on the need for intentionality in deceptive responding. Most

of these response styles for favorable self-presentation can be described as using one of three approaches: (a) providing an enhanced presentation of positive attributes, (b) denying deficits, or (c) combining these two methods. Paulhus (1984, 1998), Lanyon (2001, 2004), and Rogers (2008a) have presented varying models of favorable self-presentation (see Table 1). While none of these three models is entirely unique, each adds a different perspective to the concept of favorable self-presentation.

Paulhus' Model

More than two decades ago, Paulhus offered an original and still popular model of favorable self-presentation. Based on his factor analysis of six self-report measures of socially desirable responding, Paulhus (1984) identified two primary factors: impression management and self-deception. He defined impression management as the intentional misrepresentation of the self on overt behaviors, or consciously changing answers to present a more socially desirable façade. According to Paulhus, impression management is situationally-driven, whereas self-deception is conceptualized as a trait-like misrepresentation of the self. Individuals engaging in self-deception unintentionally, and Paulhus argues unconsciously, provide overly positive self-reports. As a result of his initial factor analysis, Paulhus created two scales to measure these styles of favorable self-presentation. These scales were similar to two of the inventories used in his original factor analysis: Sackheim and Gur's unpublished (a) Self-Deception Questionnaire and (b) Other-Deception Questionnaire, re-named as "Impression Management" (IM).

Table 1

Models of Favorable Self-Presentation

	Deliberate responding	Self-enhancement/Denial of deficits	Intended population
Paulhus			
Impression Management	Yes	Both	General
Self-deception	No	Both	General
Self-deceptive Enhancement	No	Self-enhancement	General
Self-deceptive Denial	Variable	Denial of deficits	General
Lanyon			
Extreme Virtue	Yes	Both	Forensic
Mental Health Adjustment ^a	Variable	Self-enhancement	Forensic
Denial of Psychopathology	Yes	Denial of deficits	Clinical
Self-deceptive Positivity	No	Self-enhancement	Clinical
Rogers			
Defensiveness	Yes	Denial of deficits	Clinical
Social Desirability	Variable	Both	General
Impression Management	Yes	Variable ^b	General

^aThis factor, omitted from a later model due to lack of empirical support, was subsumed within Denial of Psychopathology.

^bImpression management is an intentional attempt to present a different impression, thus it can include any nonhonest response style.

The earlier versions of Paulhus' scales, known as the Balanced Inventory of Desirable Responding (BIDR), showed significant correlation between the self-deception and impression management scales (Paulhus, 1998), suggesting the two are related or non-orthogonal concepts. In the most recent revision, Paulhus (1998) re-named the BIDR as the Paulhus Deception Scales (PDS). The PDS includes a revised self-deception scale that focuses solely on enhancement, rather than the combination of enhancement and denial of faults that previous editions had included. The rational development of the Self-Deceptive Enhancement (SDE) scale was intended to further distinguish Paulhus' two factors both conceptually and empirically. However, Salekin (2000) argued that self-deception, as it is now defined, is too similar to narcissism and arrogance to be considered a distinct response style rather than a personality trait. Besides narcissism, self-deception might also be equated with lack of insight into personality features. Despite these two limitations, Paulhus' model continues to be the most-popular, most-tested model of favorable self-presentation.

A third component of Paulhus' model incorporates the denial aspect of self-deception (Paulhus, 1999). However, it was not included in the published PDS because of strong correlations with IM and the potentially offensive content of some items. Paulhus characterized this scale, Self-Deceptive Denial (SDD), as measuring defensiveness and capturing individuals' unwillingness to admit to their psychoanalytic motivations of sex and aggression.

Lanyon's Model

Unlike Paulhus' approach with the PDS, which is a self-contained response style measure, Lanyon (2007) incorporated scales measuring favorable self-presentation into his multiscale inventory, the Psychological Screening Inventory (PSI). The PSI screens for psychological symptoms and provides several validity scales to detect favorable self-

presentation. More than any other researcher, Lanyon's contributions to the study of favorable self-presentation have spanned both strong conceptual and empirical models.

Lanyon's (2001, 2004) model has undergone several iterations during the last three decades. In 2001, Lanyon appeared to agree with Paulhus in hypothesizing that favorable self-presentation includes at least two styles. As described by Lanyon, endorsing extreme virtue and exaggerating mental health adjustment generally correspond to Paulhus' concepts of impression management and self-deception, respectively. Absent from this model is any indication regarding the motivation of respondents. Where Paulhus (1984) described impression management as intentional, Lanyon remained silent on the driving force behind endorsing extreme virtue and exaggerating mental health adjustment. These two factors are represented on the PSI as the Endorsement of Extreme Virtue (EEV) and Endorsement of Superior Adjustment (ESA) scales.

Though ESA makes strong conceptual sense, it was not empirically supported in Lanyon's (2001) factor analysis of MMPI-2, PSI, and BIDR validity scales. Using protocols from forensic clients for whom he had performed evaluations (child custody, personal injury, sex offenses, and other criminal offenses), Lanyon chose 15 validity scales to capture multiple dimensions of favorable and unfavorable self-presentation. Of the eight related to favorable self-presentation, he claimed that two measured exaggeration of mental health adjustment, four measured claims of extreme virtue, and the remaining two measured both favorable and unfavorable self-presentation on a dimensional scale. After the factor analysis failed to support ESA, Lanyon (2001, 2004) candidly observed that most items he had originally identified as describing exaggerated mental health adjustment were actually focused on the presence of psychopathology. Thus, denial of psychopathology is likely a more common response style in the context of psychological assessment, where most questions use psychopathology-anchored

questions.

In his revised model, Lanyon (2004) replaced exaggeration of mental health with its pathology-focused corollary, specifically, the denial of psychopathology. In addition to altering the second factor, Lanyon (2004) suggested a third factor: self-deceptive positivity. This factor appears to be conceptually equivalent to Paulhus' SDE, with self-deceptive positivity defined as over-claiming or overconfidence. Similar to narcissism, it is likely a stable personality trait that is prone to lack of insight (Lanyon, 2004). Despite including self-deceptive positivity in his conceptual model, Lanyon has not added a third scale corresponding to this third factor to the PSI. Thus, it appears that Lanyon's model closely maps onto Paulhus' framework, with endorsement of extreme virtue corresponding to impression management, exaggeration of mental health adjustment corresponding to self-deceptive enhancement, and self-deceptive positivity corresponding to self-deceptive denial.

Rogers' Model

While the previous two models are based on the authors' personal research and measures, Rogers (2008a) presented an overarching model of favorable self-presentation not limited to any particular measure. This model encompasses three distinct response styles: socially desirable responding, impression management, and defensiveness. Social desirability and impression management refer to broad styles of simulated adjustment and are not limited to psychological issues (Rogers, 2008a). Social desirability, similar to the Paulhus model, is a pervasive tendency for an individual to deny negative characteristics and endorse positive ones. According to Rogers, impression management is more situationally-driven and requires deliberate effort. In addition, its goal is not necessarily prosocial as it was conceptualized in Paulhus' and Lanyon's models. Accordingly, Rogers' conceptualization of impression management might best be

termed “faking for a desired objective” rather than “faking good” because it is the only response style to consider the portrayal of less socially desirable façades. For example, an individual pulled over for speeding might attempt to “play dumb” when questioned by the police officer. Under Rogers’ definition, the primary characteristic distinguishing impression management from social desirability is situation-specific motivation.

Similar to Lanyon’s denial of psychopathology and Paulhus’ self-deceptive denial, Rogers’ (2008a) construct of defensiveness includes the intentional denial or minimization of psychological difficulties. In contrast, Meehl and Hathaway (1946) used the term defensiveness to represent a test-attitude without any implication as to the consciousness or deliberateness. Assessing defensiveness in the context of a psychological evaluation is nearly as important as assessing patients’ clinical presentations. Without knowing the extent of clients’ defensiveness, clinicians can become stymied in trying to accurately understand their clinical presentation. Defensiveness is especially likely to occur when the symptoms being assessed are negatively viewed by others, such as many personality disorder symptoms. For Axis II disorders, defensiveness is the response style most relevant to clinical assessment.

Personality Disorder Assessment and Favorable Self-Presentation

Clearly, the impact of stigmatization, lack of insight, and favorable self-presentation significantly hamper attempts at accurate assessment of Axis II symptomatology (Widiger & Samuel, 2005; Westen, 1997). Based on the greater reliability of standardized measures, Widiger and Samuel (2005) recommend using both self-reports and interviews. Their strategy for accurate assessment of personality disorders is twofold: (a) administering a self-report inventory to identify possible areas of maladaptive traits and (b) following up with a semistructured interview focusing selectively on potential areas identified by the inventory. This approach

matches the suggested practice in the SCID-II manual (First, Gibbon, Spitzer, Williams, & Benjamin, 1997). It is designed to maximize diagnostic reliability while adhering to the time constraints of clinical practice. However, the authors recognized a flaw in this approach due to the effects of inadequate or inaccurate information on self-report inventories and semistructured interviews. Within the context of inaccurate information Widiger and Samuel (2005) neglected to consider the effects of favorable self-presentation. Without accurate validity scales for distorted responding, this approach does not address the realistic possibility that examinees could avoid detection altogether by minimizing their responses on the self-report inventory and not producing the necessary elevations to qualify for closer examination with the semistructured interview.

The exact prevalence of such positive distortions is unknown in clinical populations, although Rogers (2008b) suggested a higher prevalence of defensiveness than malingering and other forms of misrepresentation in forensic contexts. Unfortunately, identifying defensive responders is not a strength of diagnostic measures (see “Defensiveness on the MCMI” section for more extended discussion). Creation of effective detection strategies within diagnostic measures would assist clinicians in identifying nonhonest response styles, thereby streamlining the diagnostic process while making it more accurate and reliable.

For the diagnosis of personality disorders, the *DSM-IV-TR* (APA, 2000) suggests utilizing more than one interview over time to assess for stability of the maladaptive traits. Westen (1997) and Zimmerman and colleagues (2005) noted that most clinicians prefer to use the longitudinal, observation-based method rather than one-time interviews that ask direct questions for diagnosing personality disorders. The *DSM-IV-TR* also suggests incorporating collateral information since the individual may not view their traits as problematic or may be

unwilling to disclose them. Clinical judgment is needed in the evaluation of discrepant client versus informant data and in balancing differences between reported and observed characteristics. Both of these methods may help to address the effects of favorable self-presentation.

The most common form of diagnostic assessment in clinical practice is the unstructured interview. Despite the advantages of standardized diagnostic tools, such as semistructured interviews and self-report inventories, unstructured interviews are often used due to lack of time for comprehensive testing (Widiger & Samuel, 2005). As Zimmerman and Mattia (1999) noted, however, clinicians often miss comorbid disorders when relying only on unstructured interviews.

Prevalence

Stigmatization, lack of insight, and favorable response styles all affect the assessment of personality disorders. As such, most prevalence studies will underestimate the true rate of personality disorders. With these added challenges, most epidemiological research on personality disorders has been limited by their design and assessment methods. In particular, many studies have limited Axis II coverage, low generalizability to other clinical samples, and nonstandardized or nonvalidated methods of diagnoses. As it now stands, most epidemiological research suggests that the prevalence of personality disorders in the general population is approximately 10-15% (Grant et al., 2004; Lenzenwenger, Lane, Loranger, & Kessler, 2007).

Beyond epidemiological studies, the prevalence of personality disorders in the clinical population is most helpful because of their association with (a) the duration, recurrence, and treatment outcome of Axis I disorders and (b) their influence on treatment recommendations (Zimmerman et al., 2005). Using the recommended administration method described above, Verheul and colleagues (2000) found that the SCID-II identified at least one *DSM-III-R* Axis II

disorder in 57.0% (211 of 370) of substance users in treatment. With the Structured Interview for *DSM-IV* Personality (SIDP-IV; Pfohl, Blum, & Zimmerman, 1997), Zimmerman and colleagues (2005) determined that nearly half (45.5%, or 391 of 859) of psychiatric outpatients met diagnostic criteria for at least one of the 10 standard personality disorders or Personality Disorder Not Otherwise Specified. Moreover, Axis II comorbidity is common among outpatients diagnosed with a specific personality disorder, 60.4% warranted two or more Axis II disorders (Zimmerman et al., 2005). Therefore, the clinical impact of Axis II cannot be ignored; approximately half of all psychiatric outpatients likely experience significant impairment or distress resulting from maladaptive personality characteristics.

The large clinical prevalence estimates provided by Verheul et al. (2000) and Zimmerman et al. (2005) underscore the need for accurate assessment of Axis II pathology. Failing to disclose significant symptoms, intentionally or not, is unquestionably influential in the assessment of clinical prevalence of personality disorders. While a plethora of studies have examined the effect of favorable self-presentation on general personality measures, such as the Five Factor Model or multiscale inventories, the effect of defensiveness on the measurement of personality disorder symptomatology has received very little attention. Knowing how these response styles affect personality disorder measures is an important step toward accurate assessment and specialized treatment for these patients.

Validity Scales

Because personality disorders can alter the presentation and course of Axis I disorders and influence treatment decisions, accurate assessment of Axis II symptomatology is essential (Bakken et al., 2007; Cacciola et al., 2001; Hayward et al., 2006; Zimmerman et al., 2005). Much research has focused on integrating self and informant reports (Klein, 2003) in the

diagnosis of personality disorders to compensate for lack of insight (see Farmer, 2000, for a summary). As previously noted, relatively little research has delved into defensiveness or minimization of personality disorders.

The assessment of personality disorders via interviews and other measures relies primarily on clients' honest and accurate reporting. Stigma, lack of insight, and defensiveness could each affect responding in such a way as to provide an inaccurate assessment of personality disorders and lead to missed diagnoses (Meehl & Hathaway, 1946; Westen, 1997; Widiger & Samuel, 2005; Zimmerman et al., 2005). Because most methods, such as self-reports, rely on direct questioning of Axis II symptoms, they are easily susceptible to any type of distorted responding. As an attempt to decrease the influence of stigma and lack of insight, both self-reports and standardized interviews have been developed to collect data from different sources, including some combination of the client, the clinician, and informants. However, relatively few diagnostic personality disorder measures have validity scales to assess for distorted responding.

Validity scales have proven useful in clinical assessment to help determine examinees' response styles. Those using empirically validated detection strategies aimed at specific response styles, such as defensiveness, are likely to be most effective. Unfortunately, the lack of standardized terminology in operationalizing favorable self-presentation makes it difficult to produce validity scales that focus exclusively on one specific detection strategy. In fact, many scales purport to measure one response style but actually utilize detection strategies best suited for capturing other styles of responding (Sellbom & Bagby, 2008). Many validity scales also blend detection strategies, not allowing for the identification of one particular response style.

Rogers (2008b, p. 16) outlined the criteria necessary for sound detection strategies: "A detection strategy is a standardized method, which is conceptually-based and empirically-

validated, for systematically differentiating a specific response style (e.g., malingering or defensiveness) from other response styles (e.g., honest responding).” Detection strategies for evaluating favorable self-presentation have tended to focus on *plausible presentations*, these presentations attempt to identify defensive responders via unusual magnitudes or frequencies of characteristics commonly found in the target population. For example, defensive responders might completely deny undesirable but common behaviors, overrespond to desirable traits, or both. Unlike malingering research, few, if any, detection strategies have sought out characteristics that are atypical in the target population. For defensiveness, the rare combinations strategy, or the unlikely co-occurrence of two or more symptoms, could be reversed to capture the denial of two or more common personality disorder symptoms. This approach has been termed “denial of patient characteristics” (Rogers, 2008b).

Rogers (2008b) identified five specific detection strategies that each attempt to capture an aspect of defensiveness as it might appear on a psychological measure: denial of personal faults, affirmation of virtuous behavior blended with denial of personal faults, spurious patterns of simulated adjustment, denial of patient characteristics, and social desirability. As noted previously, research on validity scales is difficult to interpret because scales often claim to use detection strategies that they do not actually use or they blend multiple strategies. Of particular importance to this thesis are the detection strategies used on the IM, SDE, and SDD scales of the PDS (Paulhus, 1998, 1999).

Paulhus’ (1998) IM scale appears to capture denial of personal faults. Questions about nonclinical attitudes and behaviors are assessed, such as the respondent’s safety when driving above the speed limit and whether the respondent’s first impressions of people turn out to be correct. The supplementary SDE scale (Paulhus, 1999) also uses denial of personal faults, but the

focus of these questions is more intrusive and personal. Questions on this scale generally focus on negative feelings and behaviors, such as shame, anger, and cruelty. In contrast to the other two, the SDE scale (Paulhus, 1998), utilizes affirmation of virtuous behavior blended with denial of personal faults. These questions, for example, ask whether the respondent litters in the street or tries to avoid eavesdropping.

Despite the relative success of the PDS as a stand-alone measure of favorable self-presentation, validity scales built into existing measures are likely to be most beneficial to clinicians. Furthermore, because the PDS is not a measure of pathology, none of its items utilize the “denial of patient characteristics” detection strategy, which is most akin to the defensive response style. This strategy, as noted above, is inversely related to the rare symptoms strategy used in malingering detection. Creating effective validity indices on personality measures, however, has proven rather challenging. The L and K scales on the MMPI-2, two of the most popular faking good scales, are not reliable predictors of faking good (Bagby, Buis, & Nicholson, 1995; Bagby, Rogers, Nicholson, Buis, Seeman, & Rector, 1997). Bagby, Gillis, Toner, and Goldberg (1991) found that faking good and faking bad validity scales on the MCMI-II, intended to detect separate styles of responding, worked interchangeably with low faking bad scores predicting the presence of faking good and vice versa. Compounding this difficulty, response style research in personality assessment has primarily focused on general personality inventories (see Viswesvaran & Ones, 2001, for a meta-analysis), not diagnostic tools.

No Axis II diagnostic interviews include validity scales. One possible explanation for this phenomenon is that researchers believe distorted responding is more difficult to achieve during interviews. In partial support of this assumption, self-report inventories tend to overdiagnose personality disorders when compared with diagnostic interviews (Whyte, Fox, & Coxell, 2006).

However, the lower rate of diagnosis through interviews in no way lessens the need for accuracy or the inclusion of validity indices.

One useful purpose of self-report measures, as recommended by Widiger and Samuel (2005), is being used as screens prior to more in-depth assessment. Especially with screens, identifying defensiveness is a critical issue in ensuring that persons with Axis II pathology are appropriately recognized. Few self-report measures provide sufficient diagnostic assessment of *DSM-IV-TR* personality disorders, and those that do fail to adequately assess dissimulation. At this time, two Axis II diagnostic self-reports have been examined in simulation studies. The issues of most relevance to this thesis are the measures' effectiveness under defensive conditions as determined by differences in Axis II presentation between honest and defensive conditions. Noting the dearth of this type of study in the literature, Bagby and Pajouhandeh (1997) examined the susceptibility of the PDQ-4 to faking good by university students and found significant decreases on paranoid, borderline, histrionic, avoidant, and dependent scales. While the validity scales were significantly increased among fake good responders, scores on validity scales were not accurate in detecting fake good respondents. These changes on a diagnostic measure are consistent with Viswesvaran and Ones' (1999) findings that the Big 5 personality domains and social desirability scales on multiscale inventories are also highly susceptible to faking good. More research has examined dissimulation on the MCMI, described in the section below.

Defensiveness on the MCMI

The Millon Clinical Multiaxial Inventory (MCMI), a self-report personality measure now in its third edition (MCMI-III; Millon, Davis, & Millon, 1997), is distinguished from other inventories by its twin objectives to assess both Axis II diagnoses and response styles. To achieve brevity, some compromises were required. Of interest to response styles, most of the

items on validity indices are also used as indicators of clinical symptoms. As a result, the scales primarily utilize overendorsement or denial of patient characteristics as the primary detection strategies.

Two of the MCMI's modifying indices utilize denial of patient characteristics as part of a bidirectional approach to capture both favorable and unfavorable response styles. The Disclosure Index, or Scale X, is designed to measure responses that vary from the normative midrange on a composite of several scales (Millon et al., 1997). Because the Disclosure Index encompasses a broad range of clinical symptoms, the defensive detection strategy capitalizes on indiscriminate denial of patient characteristics via unexpectedly low scores. Conversely, high scorers on the Disclosure index are likely evidencing indiscriminate endorsement of patient characteristics. The scale used to detect malingering, Scale Z or the Debasement Index, utilizes a rare symptoms strategy for high scorers. Though designed to detect malingering, Sellbom and Bagby (2008) have advocated the bidirectional use of the Debasement Index. For defensiveness, very low scores on the Debasement Index can also be described as the unrealistic denial of patient characteristics or personal faults.

The MCMI-III's Desirability Index (Scale Y), the designated measure of favorable self-presentation, claims to assess whether respondents were trying to appear "socially attractive, morally virtuous, or emotionally well composed" (Millon et al., 1997, p. 128). The lack of unique Scale Y items strongly suggests that this scale utilizes denial of patient characteristics as its primary detection strategy. From a different perspective, Sellbom and Bagby (2008) described Scale Y as blending affirmation of virtuous behavior and denial of patient characteristics.

Cut scores on validity scales must strike a balance between misclassifying genuine patients as dissimulating (false positives) and not identifying dissimulators (false negatives). In

examining different uses of Desirability Index cut scores and scale configurations across four studies (two with students, one with substance users, and one with clinical outpatients), Sellbom and Bagby (2008) found a mean sensitivity of .60 for the MCMI-II and MCMI-III in classifying defensive responding. Thus, the Desirability Index misses 40% of defensive respondents (false negative rate = $1 - \text{sensitivity}$). This estimate bodes poorly for the MCMI's clinical utility in assessing highly stigmatized Axis II characteristics. In general, Sellbom and Bagby concluded that the MCMI modifying indices lack adequate support to make definitive response style classifications.

The MCMI should be lauded for its systematic attempts to capture response styles, including defensiveness. However, the application of standard cut scores in a clinical population results in unacceptable miss rates for overly favorable self-presentation. Furthermore, while several studies have examined the utility of MCMI modifying indices, few have described the effects of favorable responding on specific MCMI clinical scales. Fals-Stewart (1995; also described in Sellbom & Bagby, 2008) and Retzlaff, Sheehan, and Fiel (1991) examined its effects on the MCMI-II (see Table 2, on the next page).

Retzlaff and colleagues (1991) examined various response styles on the clinical scales of the MCMI-II by asking university students to respond under one of five sets of instructions. Participants in the defensive condition were instructed as follows: "Please take the test as if you were committed to a psychiatric hospital and wanted to be released" (Retzlaff et al., 1991, p. 468). Despite challenging instructions that require a complex, two-step simulation, students demonstrated significant shifts in Axis II scale scores between honest and defensive conditions (mean absolute $d = 0.72$). As expected, the Desirability scale increased greatly under defensive instructions. Simulators in the defensive condition had an average absolute change of $d = 1.41$

across all three validity scales.

Fals-Stewart (1995) systematically assessed the effects of defensive responding on the MCMI-II's clinical scales with a substance-dependent population. Though the defensive group was only instructed to deny substance use problems, their scores on eight of the 13 personality disorder scales were significantly lower than honest respondents' scores (mean absolute $d = 0.72$). It is likely that both the nature of substance use denial and the types of scenarios (e.g., workplace, child custody) led respondents to also conceal maladaptive personality traits. Again, defensive respondents had higher elevations on Scale Y and sizeable changes on all validity indices (mean absolute $d = 0.61$). In fact, only 12 (19.4%) of the 62 defensive respondents were unsuccessful and exceeded this cutoff for detecting fake good responding on Scale Y, the Desirability Index.

Both Retzlaff et al.'s (1991) student sample and Fals-Stewart's (1995) substance using sample showed clinically significant decreases on most personality disorder scales. These studies varied substantially in magnitude of change on most scales; nonetheless changes on two scales stand out. In particular, the Antisocial and Borderline scales—arguably containing the most interpersonally problematic Axis II symptoms—evidenced large to very large decreases (see Table 2) in the defensive conditions. This is particularly troublesome in clinical populations because, as Zimmerman and colleagues (2005) noted, knowledge of Axis II pathology helps to target interventions and increase the likelihood of success.

Table 2

Defensive Responding Effect Sizes on MCMI-II Validity and Axis II Diagnostic Scales

Validity Scale	Retzlaff et al. ^a	Fals-Stewart ^b
Desirability – Y	1.67	0.57
Disclosure – X	-0.98	-0.85
Debasement – Z	-1.57	-0.40
Validity Scales: <i>M</i> Cohen's <i>d</i> (Absolute Value)	1.41	0.61
Diagnostic Scale	Retzlaff et al. ^a	Fals-Stewart ^b
Schizoid	-0.70	-0.36
Avoidant	-1.26	-0.63
Dependent	0.12	0.10
Histrionic	-0.25	-0.74
Narcissistic	0.00	-0.84
Antisocial	-0.85	-1.24
Obsessive-Compulsive	0.92	0.93
Schizotypal	-1.14	-0.27
Borderline	-1.63	-1.00
Paranoid	-0.29	-1.06
Diagnostic Scales: <i>M</i> Cohen's <i>d</i> (Absolute Value)	0.72	0.72

Note. Cohen's *d* values were computed so negative values indicate defensive score decreases and positive values indicate defensive score increases as compared to honest scores.

^aRetzlaff, Sheehan, & Fiel (1991). This study used a between-subjects design, with samples consisting of 50 university students each in honest and defensive conditions; *d* computed using raw scores.

^bFals-Stewart (1995). This study used a between-subjects design, with samples consisting of 62 substance-users each in honest and defensive conditions; *d* computed using base rates.

Perhaps because of the general instructions provided to participants (Fals-Stewart, 1995; Retzlaff et al, 1991), it appears that defensive responders did not suppress all Axis II traits. Instead, defensive responders appear to have exaggerated characteristics on the Obsessive-Compulsive and Dependent personality disorder scales. This mixed pattern of denying and accentuating particular traits is indicative of Rogers' (2008a) response style of impression management. Scale score increases among defensive responders suggest that participants hold more favorable opinions of these personality characteristics. For example, defensive responders might endorse symptoms such as perfectionism when trying to present a favorable impression. That respondents utilize mixed patterns in any given format underscores the need for accurate detection strategies in the identification of favorable self-presentation.

Current Study

MCMI-II defensiveness studies (Fals-Stewart, 1995; Retzlaff et al., 1991) provide a conceptual basis for the current study, which examines the effects of defensiveness on personality disorder measures in an inpatient population with substance abuse and comorbid diagnoses. As recommended by Widiger and Samuel (2005), the current study uses both a self-report and a standardized interview to assess Axis II pathology. Respectively, these measures are the Structured Clinical Interview for DSM-IV - Axis II - Personality Questionnaire (SCID-II-PQ; First et al., 1997) and the Structured Interview for DSM-IV Personality (SIDP-IV; Pfohl et al., 1997). Because neither of the diagnostic measures contains validity scales, the Paulhus Deception Scales (PDS; Paulhus, 1998) is used as general measure of defensiveness.

The primary purpose of this study is to explore whether inpatients are able to deny symptoms of personality disorders on either self-report or interview-based measures. Using the SIDP-IV, honest and defensive responses are compared on both presence and severity of

symptoms. Additionally, this study examines the levels of self-favorable presentation per Paulhus' model and whether the PDS can adequately identify defensiveness on Axis II symptomatology.

Research Questions and Hypotheses

Research Question 1

Defensive responders are expected to report significantly less Axis II symptomatology than they did in the honest condition, with potentially different effect sizes between the SCID-II-PQ and the SIDP-IV. Research has consistently shown that defensive responders can alter their presentation on general personality measures (see, for example, Viswesvaran & Ones, 1999), but few studies (Fals-Stewart, 1995; Retzlaff et al., 1991) have examined the effects of defensiveness on personality disorder symptoms.

- *Hypothesis 1:* Individuals will report fewer symptoms in the defensive than honest condition, measured as the number of reported personality disorder symptoms on the SCID-II-PQ and the SIDP-IV.
- *Hypothesis 2:* A larger effect size will be observed between honest and defensive conditions on the face-valid SCID-II-PQ than the interview-based SIDP-IV.
- *Hypothesis 3:* Individuals in the defensive condition will demonstrate less severity of their Axis II pathology than in the honest condition, as measured by a total score on the SIDP-IV.

Research Question 2

Defensive responders are expected to report significantly higher levels of favorable self-presentation than they did in the honest condition, with larger differences on intentional than unintentional response styles. Based on Paulhus' theory (1984, 1998, 1999), the IM and SDD

scales would be expected to increase with simulation while the SDE scale would not be expected to change. However, Pauls and Crost (2004) demonstrated that the SDE scale also increases with simulation.

- *Hypothesis 4:* Comparing participants' PDS scale scores across honest and defensive conditions, all three scale scores will show increases in the defensive condition.

Research Question 3

What methods are most effective at identifying defensiveness in a dually diagnosed population? Because the PDS (Paulhus, 1998) lacks specific cut scores for a clinical population, alternative cut scores for this dually diagnosed population will be established to maximize both sensitivity and positive predictive power. Balancing sensitivity ($< .50$; Rogers, 2008b) and positive predictive power ($< .75$; Rogers, 2008b) will ensure that defensive responders are identified by the test and that those classified as defensive responders are, in fact, defensive responders.

Research Question 4

Are specific diagnoses or symptom clusters more or less vulnerable to defensiveness? Personality disorders are often considered constellations of rigid and inflexible traits. However, symptoms that tend to cause the most interpersonal disruption, including those classified under Borderline and Antisocial diagnoses, appear to be highly susceptible to intentional minimization. Furthermore, the MCMI-II studies (Fals-Stewart, 1995; Retzlaff et al., 1991) demonstrated that symptoms associated with greater social desirability, such as perfectionism and seeking assistance from others, can lead to increases on Obsessive-Compulsive and possibly Dependent personality disorder scales under defensive instructions.

Supplementary Question 1

Are narcissistic personality traits and self-deceptive enhancement related? Salekin (2000) and Pauls and Crost (2004), among others, have questioned whether the SDE scale, a measure of unintentional favorable self-presentation, is conflated with narcissism.

Supplementary Question 2

Do participants' reported levels of effort and perceived success predict success at reducing symptomatology? This question is intended to measure participants' insight into their ability to present more favorable impressions. Questions asked during the debriefing will address participants' reported level of effort and perceived success at obtaining the desired outcome in the defensive condition (i.e., getting admitted to the more desirable inpatient unit).

CHAPTER 2

METHODS

Design

This simulation study used a within-subjects design to examine the effect of defensive responding on personality measures. Every participant completed the same three measures under both honest and defensive instructions, allowing for each person to be used as their own comparison, thus reducing the effect of individual differences and maximizing statistical power. All data collection was completed by four graduate students with extensive training in assessment and structured interviews. The SCID-II-PQ and SIDP-IV were used to compare presence of personality disorder symptoms. In addition, the SIDP-IV results were assessed dimensionally, looking at severity of Axis II pathology. Differences in PDS scale scores were also examined to determine the effects of defensiveness on a measure of favorable self-presentation.

Though existing literature has made great strides in examining favorable self-presentations, several methodological considerations were addressed for the first time in the current study. First, previous simulation designs for Axis II disorders have used community, workplace, or college samples, but they have not sought to investigate patients with comorbid mental disorders (e.g., Holden, Starzyk, McLeod, & Edwards, 2000; Li & Bagger, 2007; Pauls & Crost, 2004). Only one study was found that utilized a clinical population; Fals-Stewart's (1995) examination of the MCMI-II in substance users. Second, unlike previous simulation studies that used between-subjects designs, this study employs a within-subjects design, described by Viswesvaran and Ones (1999) as maximizing power and reducing group differences in simulation studies. Finally, previous researchers have largely focused on self-report measures of

personality traits. Thus, this design builds upon existing literature by using a dually diagnosed clinical sample, maximizing statistical power, and using both self-report and interview-based diagnostic measures.

Participants

Participants were adult patients on the inpatient dual diagnosis unit at Timberlawn Mental Health System in Dallas, Texas. Timberlawn's Dual Diagnosis Program provides integrated treatment of both addiction and mental disorders. Group and individual psychotherapy, psychiatric services, and support services are provided. Participants were excluded if they were not able to complete the two-part study due to scheduling conflicts. All participants were fluent in English.

Research and Administrative Approval

This research project was administratively approved by the Timberlawn Mental Health System's executive board (see Appendix A) and received ethical approval from the University of North Texas Institutional Review Board (see Appendix B).

Measures

Structured Interview for DSM-IV Personality (SIDP-IV)

The SIDP-IV (Pfohl et al., 1997) is a semi-structured diagnostic interview that evaluates for *DSM-IV-TR* personality disorders with behaviors grouped by ten life themes, such as social relationships or work, rather than by diagnoses. Because of the focus on life areas rather than symptom clusters, this thematic organization of clinical inquiries may increase the likelihood that respondents will provide accurate information by decreasing face validity. A four-level rating system indicates whether a symptom is absent (0), minimally present (1), generally present for the majority of time during the last five years (2), and present plus causing distress or

impairment (3). The number of symptoms rated as present (2s and 3s) determines qualification for each particular personality disorder. As noted, the current study focused on Axis II symptoms, rather than diagnoses.

Three studies found moderate to excellent interrater reliability for the SIDP-IV. Miller, Bagby, Pilkonis, Reynolds, and Lynam (2005) found strong interrater reliability using the SIDP-IV with a clinical population (mean ICC = .90; range from .77 for Schizotypal Personality Disorder to .96 for Borderline and Avoidant Personality Disorders). Also in a clinical population, Zimmerman and colleagues (2005), demonstrated good to excellent interrater reliability for each individual personality disorder and at the cluster-level (κ s ranging from 0.61 to 1.00). Using a sample of non-treatment seeking military recruits, Jane, Pagan, Turkheimer, Fiedler, and Oltmanns (2006) examined interrater reliability for four different measurement levels on the SIDP-IV: diagnosis, number of symptoms, continuous total score, and criteria-level. Despite low base rates, their reliability estimates for number of symptoms and total score for each disorder were very good (median ICC = .80, range from .65 for Schizotypal–number of symptoms to .93 for Avoidant–total score). Of particular importance to this study, Jane and colleagues (2006) found lower reliability at the symptom-level for Axis II criteria that (a) lacked clear behavioral markers, (b) were likely to be deemed socially desirable, and (c) required insight.

Structured Clinical Interview for DSM-IV, Axis II, Personality Questionnaire (SCID-II-PQ)

The SCID-II-PQ (First et al., 1997) is a 119-item self-report that screens for the ten *DSM-IV* personality disorders and two personality disorders listed in *DSM-IV-TR* Appendix B (i.e., Depressive Personality Disorder and Passive-Aggressive Personality Disorder). The SCID-II-PQ is intended to be used prior to administering the SCID-II (First et al., 1997), a semi-structured interview that fully assesses for those 12 personality disorders. The SCID-II-PQ addresses

symptom presence (i.e., “yes” and “no” response options).

The SCID-II-PQ maximizes sensitivity and is most effective as a screening tool intended to produce high false positive rates. In terms of internal consistency, Miller et al. (2005) found a median coefficient alpha of .69 for the *DSM-IV* version, but alphas for diagnoses ranged widely from .32 for Obsessive-Compulsive to .84 for Borderline. Although psychometric data on the *DSM-IV* version is limited, the closely related *DSM-III-R* version of the SCID-II-PQ showed low false negative rates when compared with another self-report (First et al., 1997). For the purposes of the current study, the SCID-II-PQ will be used solely as a measure of personality disorder symptoms, not for diagnostic purposes.

Paulhus Deception Scales (PDS)

The PDS (Paulhus, 1998) is a self-report questionnaire designed to capture favorable self-presentation according to Paulhus’ model. The questionnaire is comprised of two 20-item scales: the Impression Management (IM) scale and the Self-Deceptive Enhancement (SDE) scale. Its 5-point ratings are reduced to dichotomous scoring (i.e., only extreme ratings count toward scale scores). The PDS manual (Paulhus, 1998) offers norms for the general population, college students, military recruits, and newly incarcerated prison inmates; and interpretive guidelines based on T score conversions for the general population and inmates. Paulhus also provided raw IM cut scores based on “general adult and general psychiatric patient populations” designed to optimize sensitivity and positive predictive power, thus minimizing misses and false alarms (Paulhus, 1998, p. 10).

The PDS has evidenced at least moderate reliability and validity across varying populations. Paulhus (1998) reported relatively strong reliability coefficients for both IM and SDE (.81 to .86 for IM and .70 to .75 for SDE). Li and Bagger (2007) found slightly lower mean

reliability coefficients (IM = .74, SDE = .68) among studies using a previous version of the PDS. However, it is unclear whether psychometric data from previous versions generalize to the current PDS because of the switch from a 7-point dichotomous scoring system to a 5-point dichotomous scoring system and the further differentiation between IM and SDE scoring procedures. Lanyon and Carle (2007) found moderate convergent validity for the current version of the PDS when compared with validity scales found on multiscale inventories for undergraduate and forensic examinee samples.

Paulhus (1998) indicated that continuous scoring may also be used on the PDS, though norms are not provided for this method. Stöber, Dette, and Musch (2002) argued for the use of a continuous scoring system as it appears to provide better internal reliability and convergent validity than the dichotomous method, but Pauls and Crost (2004) found similar reliability for both methods. Both scoring methods will be examined in the process of identifying useful cut scores.

In addition to the two standard PDS scales, Paulhus created a third 20-item scale, the Self-Deceptive Denial (SDD) scale, which he theorized better captures defensiveness (Paulhus, 1999). Paulhus declined to include SDD in his published measure due to the highly personal nature of some statements (e.g., “I rarely have sexual fantasies.”) and its “high” correlation with the IM scale. Although Paulhus did not provide the correlation, Djikic, Peterson, and Zelazo (2005) found a negligible relationship in university students ($r = .06$). Paulhus presented the SDD scale on a 7-point Likert-type scale, but it was transformed into a 5-point scale to be consistent with the IM and SDE scales of the PDS. The SDD scale is included for the current analyses as a research scale; however, it must be interpreted with caution because it lacks definitive norms and validation.

Procedures

Participant Recruitment

Seventy-one participants were recruited from Timberlawn Mental Health System's Dual Diagnosis program between January 15, 2009 and January 10, 2010. Researchers made a recruitment announcement to the patients before their first therapy group in the morning and answered questions from interested persons. Preference was given to clients who were not currently experiencing severe substance abuse withdrawal, as determined by Timberlawn staff. Timberlawn therapists also provided approval based on their perception of a patient's need to attend group and ability to understand and follow instructions. On an individual basis, the selected persons were asked about their willingness to participate and provided with more details about the nature of the study. To gauge their level of comfort with the deception required for participation, all participants were informed that they would be asked to lie to or hide information from a researcher. Though some participants expressed initial reservations, none dropped out due to this requirement.

Informed Consent

Informed consent was obtained from persons indicating interest (see Appendix B). The consent form, which detailed the procedure of the study and the participants' rights as volunteers for research, was reviewed with each person. Confidentiality, compensation, benefits, and possible risks were also described; the consent form informed participants that they would receive \$10 remuneration for their involvement in the study. As described in the informed consent, participants allowed researchers to obtain diagnostic information from their medical records. This information was retrieved from their client charts after testing or while they were completing self-report measures (see Appendix C). All demographic information (also on

Appendix C), however, was obtained directly from participants prior to Phase I.

Phase I

All measures were first administered under honest (i.e., standard) instructions to avoid potential unwanted effects of providing deceitful responses prior to honest responses, namely that their honest responses would be biased by their distorted responses. Phase I, or the honest condition, was used to establish a baseline of personality functioning. In a brief description of the purpose of the study, participants were asked to complete the following measures as honestly and accurately as possible (see Appendix D). These instructions were read aloud then provided to the participant to refer to as needed throughout Phase I. Participants then completed the SIDP-IV, SCID-II-PQ, and PDS. The SIDP-IV was administered before the SCID-II-PQ so that responses were not influenced by the face validity of the SCID-II-PQ; while SIDP-IV questions ask about behaviors in varying situations, the SCID-II-PQ questions are more directly asking about presence or absence of symptomatology. The PDS was administered last to maintain the continuity of the personality assessments and so the questions about impression management did not unduly influence the participants' reporting about their personality issues. Upon completion of Phase I, a manipulation check was used to determine whether the participant understood and followed the directions. In particular, participants were asked to identify their comprehension of instructions, level of effort, and perceived success (see Appendix E). Though simulation studies do not typically include a manipulation check for control conditions, it was used in the current study to help demarcate the transition from honest to defensive responding and to maintain a similar structure across the two phases.

Phase II

Phase II, the defensive condition, was conducted with a second interviewer who was

masked to Phase I responses. Participants were provided with defensive instructions and a hypothetical scenario (see Appendix D). As a point of clarification, the word “defensive” was not used with participants due to possible negative connotations. Instead, the instructions were introduced in a way that encourages participants to present themselves in a more favorable manner. The scenario, intended to be familiar to this population, read as follows:

You are about to be admitted to the hospital, and you’ve already figured out that there are two very different units. Patients on the East Unit seem to have fewer problems and are given more freedom. Patients on the West Unit seem to have severe personality problems and have trouble getting along with others.

While you realize you have some personality problems and might be a better match for the West Unit, you really want to get into the East Unit. Some of the perks of the East Unit are private bedrooms and personal TVs.

Participants were asked to explain the hypothetical scenario to the interviewer to ensure comprehension before beginning the defensive condition. The SIDP-IV, SCID-II-PQ, and PDS were administered again in the same order as Phase I. Throughout administration, participants were periodically reminded to continue following their instructions. As in the honest condition, participants completed a manipulation check after completing the defensive condition and answered questions regarding their motivation, effort, and perceived success (see Appendix E).

CHAPTER 3

RESULTS

Prior to data analysis on the research questions, several initial analyses were conducted. The final sample of dually diagnosed inpatients was selected by removing individuals who did not complete the study, failed the manipulation check, or had subthreshold levels of Axis II symptomatology (described in detail in “Refinement of the Sample”). After applying these filters to the 71 original participants, the final sample consisted of 52 dually diagnosed inpatients. To assess reliability, internal consistency and interitem agreement values were examined for each measure, as well as interrater reliability for the SIDP-IV. Comparisons between inpatients who did and did not meet the minimum threshold of Axis II symptomatology were conducted to ensure the representativeness of the final sample. Before addressing the hypotheses, assumptions of the inferential analyses, described later in this chapter, were tested.

Scoring and Reliability of Research Measures

PDS Scoring and Reliability

An important methodological consideration with the PDS was the selection of the more appropriate scoring (continuous vs. dichotomous) for the current results. Previously, Pauls and Crost (2004) found similar reliability for the two scoring methods. In the current sample dichotomous scoring evidenced greater reliability for SDE than continuous scoring, which demonstrated poor internal consistency (see Table 3). For the evaluation of cut scores, dichotomous scoring was used; this method is consistent with recommendations by Paulhus (1998).

For Research Question 2, comparisons across the three PDS scales were required. Because the SDD scale does not employ dichotomous scoring (Paulhus, 1999), continuous

scoring, with appropriate caveats was employed.

Table 3

Internal Consistency of PDS

	Continuous		Dichotomous	
	α	<i>M</i> Interitem <i>r</i>	α	<i>M</i> Interitem <i>r</i>
IM	.72	.11	.67	.08
SDE	.42	.04	.70	.12
SDD	.56	.06	-- ^a	-- ^a

Note. Across scales and scoring methods, 31.5% of all interitem correlations were negative.

^aAlpha was not computed because SDD was not scored dichotomously

The lack of internal consistency seen in the current sample differs markedly from the alphas reported by Paulhus (1998) whose alphas were high for IM (.81 to .86) and moderate for SDE (.70 to .75). Understandably, the SDD alphas were not reported because it is not currently used for clinical interpretation (Paulhus, 1999). In the current study, the low average interitem correlations underscore the lack of homogeneity within scales. In fact, nearly one-third of the items were negatively correlated with other items on the same scale, even after reverse scoring of the appropriate items.

The marked discrepancy between levels of internal consistency in the current study and Paulhus' (1998) is difficult to explain. One potential explanation is that Paulhus (1998) did not differentiate between clinical and community samples when reporting his alphas. Therefore, it is possible that his clinical sample also evidenced relatively low reliabilities, but that it was obscured by adding the community sample. In addition, the lack of applicability of some PDS questions may also have decreased the reliability in the current dually-diagnosed sample. For example, the question inquiring about international travels appears inapplicable to markedly impaired inpatients that mostly relied on disability benefits for income.

Interestingly, alpha coefficients improved dramatically for two scales when administered under defensive conditions. Alphas increased to an adequate level for SDD ($\alpha = .78$) and was excellent for IM ($\alpha = .90$), but SDE ($\alpha = .41$) remained essentially unchanged. This dramatic improvement might reflect the more face-valid questions on IM and SDD. Thus, participants in the defensive condition could easily discern how to make a highly positive impression and, therefore, present a more consistent image than under the honest condition.

Reliability of Axis II Measures

Consistent with previous research, both measures of personality disorders demonstrated widely ranging values for internal consistency for individual disorders, but acceptable levels for the clusters and total score (see Table 4). The SCID-II-PQ's median alpha of .72 is comparable to Miller et al. (2005; $\alpha = .69$) and slightly better than previously found among substance-dependent inpatients ($\alpha = .63$ on the *DSM-III-R* version; see Ball, Rounsaville, Tennen, & Kranzler, 2001). On SIDP-IV scales, Miller and colleagues also found moderate reliability with a median of .72, slightly higher than that found in the current sample (median $\alpha = .58$; see Table 4). Consistent with Miller et al.'s results, the two SIDP-IV scales with the highest reliability were Avoidant Personality Disorder and Borderline Personality Disorder and the SCID-II-PQ scale with the highest reliability was Borderline Personality Disorder. In stark contrast, Schizotypal Personality Disorder manifested very poor reliability in the current study with an alpha of .21. Three of the nine criteria for Schizotypal Personality Disorder are assessed with observational criteria (higher than for any other Axis II diagnosis), and one of those items was removed because of zero variance (i.e., all participants received scores of 0 on "behavior or appearance that is odd, eccentric, or peculiar"). With a more restricted range (0-10) than any other disorder (other ranges varied from 13-22), the current sample's low heterogeneity (Streiner, 2003) of

schizotypal symptomatology contributes to the SIDP-IV's low alpha. Restricted ranges and alphas below .50 were also observed on the SIDP-IV's histrionic scale and the SCID-II-PQ's schizoid scale.

Table 4

Internal Consistency of Personality Disorder Clusters and Diagnoses for SCID-II-PQ and SIDP-IV

Scale	SCID-II-PQ		SIDP-IV	
	α	<i>M</i> Interitem <i>r</i>	α	<i>M</i> Interitem <i>r</i>
Cluster A	0.80	.13	0.69	.11
Paranoid	0.76	.28	0.57	.20
Schizoid	0.41	.09	0.58	.20
Schizotypal	0.73	.20	0.21	.07
Cluster B	0.86	.09	0.80	.10
Antisocial*	0.78	.19	0.62	.17
Borderline	0.82	.23	0.63	.16
Histrionic	0.71	.26	0.48	.09
Narcissistic	0.76	.16	0.57	.14
Cluster C	0.70	.09	0.69	.08
Avoidant	0.64	.20	0.75	.30
Dependent	0.55	.14	0.58	.16
Obsessive-Compulsive	0.55	.12	0.54	.10
Total	0.83	.09	0.84	.06

*For Antisocial Personality Disorder, the SCID-II-PQ only measures Criterion C, Conduct Disorder.

Interrater reliability was also assessed on a subsample of 10 participants via independent ratings by two graduate students with extensive training in assessment and structured interviews. As a third examiner, the interrater's scores were not used in Phase I or Phase II analyses. Item-level agreement was considered paramount in the current study because most analyses relied on item-level data rather than categorical diagnostic information. The single measure intraclass correlation coefficient assessing item-level agreement was excellent, $ICC(1,1) = .92$, 95% CI [.91, .93], $p < .001$. With excellent interrater reliability, observed differences between conditions are likely to be accurate.

Refinement of the Sample

Manipulation Check

A rigorous simulation design requires a manipulation check to ensure participants adhered to the simulation instructions. In this study, participants were periodically reminded to follow instructions. After completing each phase, a manipulation check was performed to ascertain whether instructions were accurately recalled. For Phase I, every participant had accurate recall and reported that they answered honestly. In comparison, seven participants incorrectly described their experimental instructions in Phase II (i.e., present themselves in a favorable manner). Two of these individuals also failed to meet the minimum threshold for Axis II symptomatology (see “Exclusion Criteria” below). Additionally, three participants declined to complete both conditions; due to side effects from medications, they found it difficult to maintain the alertness necessary to complete the study.

Exclusion Criteria

Participants with very few Axis II traits will have difficulty minimizing symptoms of personality disorder because of the floor effect. To avoid this, 11 participants who met diagnostic criteria for fewer than 10 *DSM-IV-TR* personality disorder symptoms were excluded from further analyses. This minimum criterion was established based on the prevalence of Personality Disorder Not Otherwise Specified (PD NOS) in a large clinical sample. Specifically, Wilberg, Hummelen, Pedersen, and Karterud (2008) found that the average number of Axis II symptoms reported was 8.8 ($SD = 3.5$, $N = 1516$). Thus, selecting 10 symptoms as the inclusion threshold made it likely that participants would, at the minimum, meet diagnostic criteria for PD NOS. More importantly, this threshold ensured that participants are experiencing considerable impairment or distress as a result of their Axis II pathology.

Representativeness of Final Sample

The current sample evidenced characteristics (see Table 5) consistent with past research (i.e., Rush & Koegl, 2008) on dually diagnosed inpatients. Prior to exclusion of inpatients with subthreshold Axis II symptomatology, most (79.0%) participants were unemployed and receiving disability benefits; these findings—while high—are still lower than the 91.8% unemployed in Rush and Koegl’s Canadian sample (enrollment in disability benefits was not reported). While both samples were primarily male (71.4% and 64.4% respectively), the current sample was slightly ($d = 0.49$) younger ($M = 43.84$, $SD = 10.35$) than Rush and Koegl ($M = 49.9$, $SD = 12.6$).

Table 5

Differences between Final Sample and Excluded Participants on Gender, Ethnicity, Marital Status, and SES

	Final Sample ($n = 52$)		Excluded ($n = 11$)		χ^2	df	p^a
	n	%	n	%			
Gender					2.48	1	.12
Male	35	77.78	10	22.22			
Female	17	94.44	1	5.56			
Ethnicity					4.08	2	.13
European American	21	87.50	3	12.50			
African American	22	73.33	8	26.67			
Other	9	100.00	0	0.00			
Marital Status					2.87	2	.14
Single	18	94.74	1	5.26			
Married	15	78.95	4	21.05			
Divorced/Widowed	19	76.00	6	24.00			
SES					0.53	1	.47
Lower	22	88.00	3	12.00			
Middle/Upper	30	81.08	7	18.92			
Employment					0.87	1	.35
No Disability	12	92.31	1	7.69			
Disability	40	81.63	9	18.37			

Note: SES and Employment missing for one person each in Excluded group. Power for these analyses was limited by some cells having fewer than 5 observed cases; to preserve power, categories were combined when applicable (e.g., Hispanic American and Asian American were combined with “Other,” and full-time, part-time, and unemployed were combined into “No Disability”).

^aUsing Bonferroni correction of α/n , or $.05/5$, p required for significance is .01.

For purposes of generalizability, participants included in the final sample were compared on demographic and clinical data to those excluded for having fewer than 10 symptoms. Gender, ethnicity, marital status, socioeconomic status (SES) and employment status were assessed with χ^2 analyses (see Table 5). Using Welch's correction,¹ one-way ANOVAs were performed with age, Global Assessment of Functioning (GAF) as rated by hospital staff at admission, and highest completed level of education as dependent variables (see Table 6). Likely due to the exclusion criterion (i.e., < 10 personality disorder symptoms), the excluded participants evidenced better psychological functioning via moderately higher GAF scores ($d = 0.62$). Overall, more than double (62.0%) of the final sample was categorized as severely impaired (GAF < 30), as compared to those excluded (27.3%), $\chi^2(1, N = 61) = 4.41, p = .04$. On all other variables, the two groups were comparable (see Tables 5 and 6) indicating no selection bias in establishing the final sample.

Table 6

Differences between Final Sample and Excluded Participants on Age, GAF, and Highest Grade Achieved

	Final Sample ($n = 52$)				Excluded ($n = 11$)				F	df	p^a	d
	M	SD	95% CI		M	SD	95% CI					
			LL	UL			LL	UL				
Age	43.06	10.50	40.08	46.04	48.00	9.78	41.43	54.57	2.36	1, 15.12	.15	0.46
GAF	26.84	6.19	25.08	28.60	30.55	4.72	27.37	33.72	4.92	1, 18.45	.04	0.62
Grade	12.43	2.56	11.70	13.16	13.00	1.41	12.05	13.95	1.06	1, 25.50	.31	0.24

Note. Welch's correction used for all ANOVAs. GAF missing values for 2 in the Included group.

^aUsing Bonferroni correction of α/n , or $.05/3$, p required for significance is $.02$.

The final sample is both representative of the broader population of dually diagnosed inpatients (Rush & Koegl, 2008) and similar to Timberlawn's other inpatients, who were

¹ Though none of the analyses violated the homogeneity of variances assumption, the highly unequal group sizes still limit the power of these tests to find significant differences (observed power ranged from .11 to .45 for the traditional ANOVAs).

excluded based on a priori criteria (see Tables 5 and 6). Of the 52 dually diagnosed inpatients, most were male (67.3%) and either African American (42.3%) or European American (40.4%). A substantial majority (82.7%) had obtained at least a high school degree or GED ($M = 12.43$, $SD = 2.56$). The final sample evidenced considerable impairment, with 76.9% receiving disability benefits and 62.0% being classified as severely impaired ($M = 26.84$, $SD = 6.19$).

Parametric Assumptions

Prior to statistical analyses of the hypotheses, the data were examined to evaluate the requisite assumptions of parametric testing. Unless otherwise noted, all assumptions were met.

Missing Data and Scoring Procedures

Two procedures were undertaken to account for missing data. All PDS protocols had five or fewer missing answers (17 in Phase I and 18 in Phase II) and, thus, were able to be adjusted using Paulhus' (1998) recommended method of pro-rating. On the SCID-II-PQ, ratios were created to reflect the proportion of "yes" responses (indicating symptom presence) out of total responses, eliminating unanswered questions from the ratio. No more than two questions were omitted by anyone in the final sample, with just three incomplete protocols in Phase I and five in Phase II. This scoring method is referred to as "reported symptomatology" in the current study. This ratio also provided a measure that was directly comparable with reported symptomatology on the SIDP-IV (counting as present items that were rated 2 or 3 and excluding clinician-rated items). No questions were omitted on the SIDP-IV.

Besides looking just at participant-reported symptomatology, the SIDP-IV was scored two additional ways that incorporated clinician-rated items. Symptom counts (i.e., number of participant-reported *and* clinician-rated symptoms scored as 2 or 3) provide a better representation of DSM-IV-TR criteria. Additionally, the SIDP-IV was scored dimensionally,

summing any rating (i.e., 1, 2, or 3) of participant-reported *and* clinician-rated symptoms. In general, dimensional ratings better represent the intensity and pervasiveness of personality disorder symptoms.

Axis II Symptomatology

The current study was predicated on dually diagnosed inpatients manifesting substantial Axis II disorders and symptomatology (Bakken et al., 2007; Cacciola et al., 2001; Damen, DeJong, & Van der Kroft, 2004; Jahng et al., 2004) so that defensive responding would not be compromised by a floor effect. Importantly, self-report measures, especially those intended as screens, are intentionally designed to overestimate disorders as part of an initial evaluation (i.e., minimize false-negatives). Therefore, the Phase I (honest) SIDP-IV was used to determine Axis II symptomatology.

Dually-diagnosed inpatients evidenced moderately high rates of personality disorder diagnoses, as expected for this clinical population. Prior to final refinement, 57.8% (i.e., 41 of 71) of the inpatients met diagnostic criteria for at least one of the 10 specific personality disorders. The prevalence rates in the current study are similar to those seen in a Dutch opioid-dependent inpatient sample (Damen et al., 2004), where 70.0% had at least one personality disorder. In contrast, just 31.4% of Zimmerman and colleagues' (2005) psychiatric outpatient sample met this standard. Of the 41 current participants with at least one personality disorder, 63.4% met diagnostic criteria for two or more personality disorders. High comorbidity on Axis II is expected when using structured interviews in clinical settings (Zimmerman & Mattia, 1999), with similar levels of comorbidity found by Zimmerman and colleagues (60.4%) and Damen and colleagues (71.4%).

The final sample manifested higher prevalence rates (see Table 7) as a result of excluding

patients without major Axis II pathology (i.e., less than 10 symptoms). Nearly two-thirds (65.4%) of the final sample qualified for at least one Axis II diagnosis. Comorbidity remained present but somewhat lower, with 42.3% of the final sample receiving two or more personality disorders. Of particular interest, fully two-thirds (67.1%) of the current sample’s diagnoses came from Cluster B (see Table 8). Overall, the final sample averaged close to 19 Axis II symptoms with a range from 10 to 41 symptoms (see Table 8). Among the 18 inpatients who did not meet diagnostic criteria for a personality disorder, their symptoms ranged from 10 to 15.

Table 7

Comorbidity of Personality Disorders among the Final Sample

	Final Sample (<i>n</i> = 52)		
	<i>n</i>	%	Cumulative %
≥ 3 Diagnoses	9	17.31	17.31
2 Diagnoses	13	25.00	42.31
1 Diagnosis	12	23.08	65.38
No Diagnosis	18	34.62	100.00

Note. Diagnoses made with SIDP-IV during Phase I of testing. Among those excluded (*n* = 11), none met criteria for a personality disorder.

The final sample exhibited very high dimensional ratings for Cluster B, primarily on Antisocial and Borderline symptomatology (see Table 8). The history of behavior related to drug use likely increases the intensity of Antisocial traits, whose symptoms include illicit and impulsive behaviors. Jahng and colleagues (2011), in a community-based epidemiological survey, found that Cluster B was more highly related to substance use disorders than the other clusters after controlling for general personality pathology. In both past and current research, shared features between Cluster B and substance use disorders, such as externalizing and disinhibition, likely contribute to the relationship. In the current sample, 55.6% of all Axis II pathology was found on Cluster B; with a combined average score of 23.09, Antisocial and

Borderline characteristics account for more than one-third (36.7%) of the dimensional pathology ratings.

Table 8

Axis II Symptomatology among the Final Sample of Dually Diagnosed Inpatients

	Dimensional				Symptoms				Diagnoses
	<i>M</i>	<i>SD</i>	95% CI		<i>M</i>	<i>SD</i>	95% CI		
			LL	UL			LL	UL	
Cluster A	11.40	6.63	9.56	13.25	3.21	2.48	2.52	3.90	6
Paranoid	5.23	3.63	4.22	6.24	1.38	1.35	1.01	1.76	2
Schizoid	3.42	3.27	2.51	4.33	1.10	1.21	0.76	1.43	4
Schizotypal	3.79	2.35	3.13	4.44	1.08	1.08	0.78	1.38	0
Cluster B	34.94	12.55	31.45	38.44	10.52	4.98	9.13	11.91	47
Antisocial	11.13	4.54	9.87	12.40	3.46	1.84	2.95	3.97	24
Borderline	11.96	4.97	10.58	13.35	3.71	2.00	3.15	4.27	16
Histrionic	5.48	3.43	4.53	6.44	1.63	1.42	1.24	2.03	2
Narcissistic	6.37	4.10	5.22	7.51	1.71	1.65	1.25	2.17	5
Cluster C	16.50	8.60	14.11	18.89	5.21	3.11	4.34	6.08	17
Avoidant	5.46	5.06	4.05	6.87	1.73	1.73	1.25	2.21	8
Dependent	4.13	3.86	3.06	5.21	1.31	1.52	0.89	1.73	1
Obsessive-Compulsive	6.90	4.05	5.78	8.03	2.17	1.49	1.76	2.59	8
Total	62.85	21.42	56.55	68.81	18.94	8.12	16.68	21.20	70

Note. SIDP-IV was used for DSM-IV-TR diagnoses and ratings during Phase I of testing. “Diagnoses” refers to the number of people who met diagnostic criteria. The SIDP-IV counts one symptom for both Schizoid and Schizotypal; it was used only once in computing Cluster A and Total values.

Defensiveness on Personality Disorder Measures

A central issue is whether individuals can successfully conceal problematic personality traits when it would be advantageous for them to do so. The first and fourth research questions, combined into one factorial ANOVA (see Tables 9 and 10, next page) and described throughout this section, explored whether dually diagnosed inpatients can conceal Axis II symptomatology and in what ways they are able to do so. As predicted in Hypothesis 1, the experimental condition produced a very large effect size (condition $\eta_p^2 = .60$). Through defensive responding, dually diagnosed inpatients greatly reduced their maladaptive personality traits on both interview-based and self-report measures (overall $d_s \geq 1.00$; see Table 10). Thus, it appears that

both the interview and the self-report are highly susceptible to defensive responding.

Table 9

Effects of Condition, Measure, and Cluster on Reported Symptomatology

Effect	<i>MS</i>	<i>df</i>	<i>F</i>	<i>p</i> ^a	<i>partial</i> η^2
Condition	3.01	1,46	67.92	< .001*	.60
Measure	3.10	1,46	81.90	< .001*	.64
Cluster	0.07	2,92	3.14	.05	.06
Measure X Condition	0.05	1,46	7.16	.01	.14
Measure X Cluster	0.21	2,92	27.63	< .001*	.38
Condition X Cluster	0.04	2,92	3.63	.03	.07
Measure X Condition X Cluster	0.07	2,92	14.05	< .001*	.23

Note. 2 X 2 X 3 Repeated-Measures ANOVA with IVs condition (honest, defensive), measure (SIDP-IV, SCID-II-PQ), and cluster (A, B, C). A ratio of yes responses (“yes” on SCID-II-PQ and ratings of 2 or 3 on SIDP-IV) over total questions was used as the dependent variable to compare reported symptomatology on the self-report and interview-based measures, which rate different numbers of questions. For symptom endorsements, the examiner’s SIDP-IV ratings were not included in this ratio. It gives an indication as to the proportion of symptoms participants endorsed on either measure.

^aUsing Bonferroni correction of α/n , or .05/7, *p* required for significance is .007.

Traditionally, measures with close-ended items (i.e., self-report measures) have been seen as more susceptible to defensive responding than measures that allow for open-ended responses (Meehl, 1945). In testing Hypothesis 2, SIDP-IV responses unexpectedly showed a slightly larger decrease ($d = 1.40$; see Table 10) than SCID-II-PQ responses ($d = 1.06$), suggesting that Axis II interviews with open-ended questions are at least as susceptible as self-reports in a true-false format (measure X condition $\eta_p^2 = .14$; see Table 9).

Self-reports are generally expected to produce higher response rates and higher levels of pathology than interviews (see Whyte, Fox, & Coxell, 2006). Consistent with its format and its use as a screen, the SCID-II-PQ has an intentionally high false-positive rate to minimize false-negatives (First et al., 1997). This method variance was supported, as the SCID-II-PQ showed much more reported symptomatology than the SIDP-IV (measure $\eta_p^2 = .64$; see Table 9). This

pattern persisted in both honest ($d = 1.15$) and defensive ($d = 0.90$) conditions. Interestingly, these differences seemed to vary by cluster (measure X cluster $\eta_p^2 = .38$) with honest SIDP-IV response rates nearly as high as SCID-II-PQ rates on Cluster B but lower on Clusters A and C (see Table 10 and Figure 1); this pattern is best explained in consideration of the three-way interaction.

Table 10

Differences between Honest and Defensive Groups on Axis II Clusters and Diagnoses

	Honest				Defensive				<i>d</i>
	<i>M</i>	<i>SD</i>	95% CI		<i>M</i>	<i>SD</i>	95% CI		
			LL	UL			LL	UL	
SIDP-IV	.25	.10	.22	.28	.11	.10	.08	.14	1.40
Cluster A	.17	.13	.13	.20	.08	.10	.05	.11	0.82
Cluster B	.31	.15	.27	.36	.12	.13	.08	.16	1.35
Cluster C	.22	.13	.18	.26	.12	.11	.09	.15	0.83
SCID-II-PQ	.39	.14	.35	.43	.23	.16	.18	.27	1.06
Cluster A	.42	.19	.37	.48	.22	.18	.16	.27	1.08
Cluster B	.37	.15	.32	.41	.21	.17	.16	.26	1.00
Cluster C	.41	.17	.36	.46	.28	.19	.22	.33	0.72

Note. See Table 9.

Defensive responding produced a differential pattern for Cluster B as opposed to the other two clusters, with responses also varying by measure (measure X condition X cluster ($\eta_p^2 = .23$; see Table 9)). Inspection at the cluster level illustrates these patterns for the exploratory fourth research question. While total changes on the SIDP-IV were greater than those on the SCID-II-PQ, closer examination (see Table 10 and Figure 1) reveals that Cluster B is primarily responsible for the SIDP-IV's larger effect size. Consistent with its higher prevalence under honest responding, Cluster B evidenced a dramatic decrease and a larger effect size ($d = 1.35$) from honest to defensive condition than those seen for Cluster A ($d = 0.82$) and Cluster C ($d = 0.83$). Moreover, the Cluster A and Cluster C effect sizes on the SIDP-IV were smaller than

those seen on the SCID-II-PQ. Thus, the SIDP-IV's greater susceptibility to defensive responding seems limited to Cluster B.

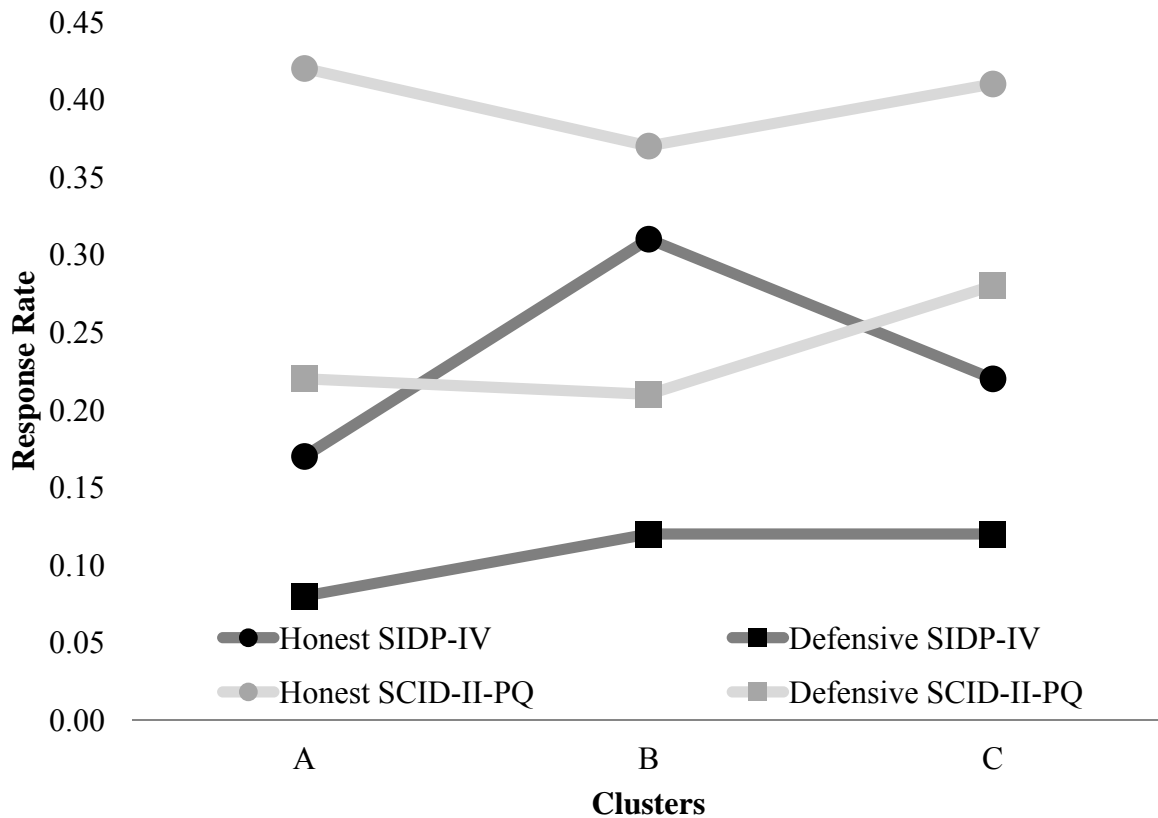


Figure 1. Proportion of reported Axis II symptomatology by condition, measure, and cluster. See Table 10 for exact response rates.

Sample characteristics and prevalence of Axis II symptoms must also be considered when examining these cluster-level changes. For example, the heightened change of Cluster B on the SIDP-IV ($d = 1.35$; see Table 10) may be due to the pervasiveness of Cluster B symptoms in the current sample as this very large difference between conditions was unmatched by other comparisons on either measure (d s range from 0.72 to 1.08). Logically, the areas with the greatest potential susceptibility to defensive responding (i.e., the most change) are those that had the highest levels of reported symptomatology under honest responding. Rather than identifying

which clusters were most susceptible on each format, this finding suggests that the clusters with the highest honest response rates (i.e., Cluster B) experienced the largest decrements across both interview and self-report measures.

Interestingly, defensive respondents endorsed a very small number of symptoms approximately equally across the three clusters, with 8-12% on the SIDP-IV and 21-28% on the SCID-II-PQ. This pattern differs from honest responding where prevalence rates differed between clusters, like patterns found in other clinical populations (Damen et al., 2004; Zimmerman et al., 2005). Across both measures, Cluster C showed the smallest change between conditions (SCID-II-PQ $d = 0.72$, SIDP-IV $d = 0.83$) and the highest Phase II levels of reported symptomatology (SCID-II-PQ rate = 28%, SIDP-IV rate = 12%), suggesting greater social acceptance of these symptoms (e.g., perfectionistic, depending on others, and hard-working) as they relate to the scenario posed in Phase II.

The SIDP-IV provides an opportunity to go beyond symptoms per se and investigate the dimensional nature and severity of maladaptive personality functioning. For Hypothesis 3, defensive responses were compared directly with the honest SIDP-IV ratings (including observational items; see Tables 8 and 11). As with the response rates described above, dimensional scores decreased more than 50% between conditions (see Tables 8 and 11; $d_{\text{total}} = 1.50$). Looking strictly at symptoms, a floor effect likely diminishes the magnitude of change for some personality disorders; nonetheless, the decrease is again greater than 50% and most effect sizes remain large ($d_{\text{total}} = 1.39$; cluster d s from 0.87 to 1.36).

Concealment of serious Axis II pathology was achieved by an impressive 65.4% of defensive responders, when using the previously established threshold (i.e., < 10 Axis II symptoms). Thus, when asked directly about their functioning, approximately two-thirds of the

inpatients successfully masked Axis II symptomatology that ought to inform clinicians' treatment decisions and recommendations.

Table 11

Axis II Symptomatology Reported under Defensive Instructions

	Dimensional					Symptoms				
	<i>M</i>	<i>SD</i>	95% CI		<i>d</i> ₁	<i>M</i>	<i>SD</i>	95% CI		<i>d</i> ₂
			LL	UL				LL	UL	
Cluster A	5.02	5.35	3.53	6.51	1.06	1.31	1.79	0.81	1.81	0.89
Paranoid	2.21	2.72	1.45	2.97	0.94	0.52	0.90	0.27	0.77	0.75
Schizoid	1.73	1.75	1.24	2.22	0.64	0.62	0.72	0.42	0.82	0.48
Schizotypal	1.40	2.08	0.82	1.98	1.08	0.29	0.67	0.10	0.47	0.88
Cluster B	15.79	12.36	12.35	19.23	1.54	4.23	4.26	3.04	5.42	1.36
Antisocial	4.33	5.29	2.85	5.80	1.38	1.21	1.78	0.72	1.71	1.24
Borderline	4.50	4.76	3.18	5.82	1.53	1.17	1.65	0.71	1.63	1.39
Histrionic	3.87	2.86	3.07	4.66	0.51	1.10	1.27	0.74	1.45	0.39
Narcissistic	3.10	3.50	2.12	4.07	0.86	0.75	1.27	0.40	1.10	0.65
Cluster C	10.08	6.78	8.19	11.97	0.83	2.79	2.39	2.12	3.45	0.87
Avoidant	2.37	3.06	1.51	3.22	0.74	0.67	1.04	0.38	0.96	0.74
Dependent	3.06	2.95	2.24	3.88	0.31	0.90	1.09	0.60	1.21	0.31
Obsessive-Compulsive	4.65	3.51	3.68	5.63	0.59	1.21	1.27	0.86	1.57	0.69
Total	30.89	21.26	24.97	36.80	1.50	8.33	7.07	6.36	10.29	1.39

Note. See Table 8. For effect sizes, *d*₁ = change in severity and *d*₂ = change in number of symptoms (see Table 8 for honest scores).

The current study demonstrates convincingly that dually diagnosed inpatients are able to significantly alter their Axis II presentation on interview-based measures. Across all three methods of evaluating SIDP-IV changes, inpatients with Axis II symptoms produced very large effect sizes ranging from 1.39 to 1.50 and dropped their scores by more than half (56.0% in reported symptomatology, 50.9% in dimensional scores, and 56.0% in total symptoms). As with response patterns on the SIDP-IV and SCID-II-PQ, certain clusters and disorders were less affected than others (see Table 11) by defensive responding. At the disorder-level, however, a floor effect is attenuating results; besides Antisocial and Borderline Personality Disorders, the other eight disorders averaged just 1-2 symptoms in Phase I, limiting the ability to reduce pathology. With a wider range, dimensional scores provide a more robust measure of change.

Those diagnoses with the smallest effect sizes (i.e., d s below 0.75) include Schizoid, Histrionic, Avoidant, Dependent, and Obsessive-Compulsive Personality Disorders. One potential explanation is that these disorders have fewer socially undesirable traits than the other five personality disorders. In fact, some of these traits—for example, depending on or getting along with others—could be viewed as positive in a social context like impressing hospital staff, as proposed by the simulation instructions.

As noted, inpatients with significant Axis II pathology were clearly able to minimize both the presence and severity of their symptoms on diagnostic measures. Concordantly, defensive responders also dramatically reduced the number of Axis II diagnoses (see Table 12). Using odds ratios (with the larger number in the numerator to aid in interpretation), participants were 9.02 times less likely to meet criteria for any Axis II disorder in the defensive than in the honest condition. Among the 65.4% with personality disorders, 73.5% (25 of 34) successfully masked their Axis II diagnoses. Not surprisingly, Antisocial Personality Disorder—with many symptoms that are viewed very pejoratively—evidenced the largest reduction of 75.0%. Individual disorders with few diagnoses in the honest condition encountered a floor effect; as a result, odds ratios could not be computed or resulted in large confidence intervals. At the cluster level, however, marked differences were observed between honest and defensive conditions with ORs ranging from 6.47 to 10.96.

Table 12

Differences between Honest and Defensive Conditions in Prevalence of Personality Disorders

	Honest		Defensive		% Decrease	OR	95% CI	
	<i>n</i>	%	<i>n</i>	%			LL	UL
No Diagnosis	18	34.62	43	82.69	-138.9	9.02	3.60	22.60
1 Diagnosis	12	23.08	6	11.54	50.0	2.30	0.79	6.69
≥ 1 Diagnosis	22	42.31	3	5.77	86.4	11.98	3.30	43.47
Cluster A	6	8.57	1	1.43	83.33	6.47	0.76	55.21
Paranoid	2	3.85	1	1.92	50.00	2.04	0.18	23.22
Schizoid	4	7.69	0	0.00	100.00	--	--	--
Schizotypal	0	0.00	0	0.00	0.00	--	--	--
Cluster B	47	67.14	11	15.71	76.60	10.96	4.85	24.74
Antisocial	24	46.15	6	11.54	75.00	6.57	2.39	18.05
Borderline	16	30.77	3	5.77	81.25	7.26	1.97	26.80
Histrionic	2	3.85	1	1.92	50.00	2.04	0.18	23.22
Narcissistic	5	9.62	1	1.92	80.00	5.43	0.61	48.15
Cluster C	17	24.29	3	4.29	82.35	7.16	1.99	25.74
Avoidant	8	15.38	1	1.92	87.50	9.27	1.12	77.07
Dependent	1	1.92	0	0.00	100.00	--	--	--
Obsessive-Compulsive	8	15.38	2	3.85	75.00	4.55	0.91	22.55
Total	70	100.00	15	21.43	78.57	--	--	--

Note: SIDP-IV was used for DSM-IV-TR diagnoses. OR could not be calculated for changes in which 0 was in the denominator. Percentages for clusters are based on total numbers of diagnoses in the honest condition (70), and percentages for diagnoses are the proportion of participants (out of 52) who received that diagnosis.

An unexpected finding involved the stable reporting or possible exaggeration of certain Axis II symptoms under the defensive condition. Eleven (21.2%) inpatients in the current study failed to reduce (i.e., did not change or even increased) at least one of the three personality disorder scoring strategies (i.e., symptom level on SCID-II-PQ, symptom level on SIDP-IV, and dimensional scoring on SIDP-IV). Of these individuals, seven (63.6%) increased scores on at least one method, while the remaining four (36.4%) decreased on two scoring methods and showed no change on the third. Despite their varied approaches, all 11 of these inpatients responded correctly to the manipulation check and, therefore, were not excluded from analyses.

A few patterns emerged when examining the characteristics of inpatients whose symptomatology did not uniformly decrease. Nearly half (45.5%) of these 11 inpatients were

among the 18 who did not meet diagnostic criteria for a personality disorder in Phase I. With little pathology to minimize, they may have undertaken different strategies (i.e., other than defensiveness, such as overendorsing virtues) to adhere to the Phase II instructions. These alternate response styles were not assessed with the current study's design. Additionally, of the six participants who retained a diagnosis of Antisocial Personality Disorder in the defensive condition (see Table 12), four of them were among the inpatients whose scores did not uniformly decrease. It is possible that these four inpatients lacked adequate insight to recognize that endorsing Antisocial and other Axis II symptoms could diminish their chances of appearing well-adjusted and being placed on the more desirable East Unit.

Defensiveness on Paulhus Deception Scales

The PDS is often used in professional settings to assess favorable self-presentation, but it has never been researched specifically with defensive responding and Axis II traits in a clinical sample. The second research question explored the effects of defensive responding in the currently dually diagnosed sample. As predicted, all three PDS scales showed very large increases² under defensive responding (condition $\eta_p^2 = .54$, $ds > 1.00$; see Tables 13 and 14). In addition to minimizing Axis II pathology, inpatients clearly altered other aspects of their presentations in Phase II. Because all three scales significantly increased under instructions to intentionally distort responses, the differential response to defensive responding (condition X scale $\eta_p^2 = .14$) appears negligible. SDE and SDD scales, designed to capture unconscious, self-deceptive response styles, were equally influenced by an intentional response style, suggesting they would

² Error in a dependent measure leads to diminished power in analyses. While the above findings are in line with expectations, it is important to note the low Cronbach's alphas (see Table 3) found for SDE and SDD when given to honest responders. A more internally consistent measure might have produced even greater magnitudes of difference across honest and defensive conditions.

have minimal ability to differentiate between response styles. However, inpatients may have utilized more than one response style in their response to PDS questions, an approach not discernible given the current study's design. Without a clear criterion for unconscious deception, it is difficult to ascertain SDE and SDD's discriminant validity.

Table 13

Effects of Condition and Scale on PDS Scores

Effect	<i>MS</i>	<i>df</i>	<i>F</i>	<i>p</i> ^a	<i>partial η</i> ²
Condition	14448.61	1, 51	58.84	< .001	.54
Scale	973.67	2, 102	7.21	.001	.12
Condition X Scale ^b	656.13	1.68, 85.88	8.43	.001	.14

Note. Continuous scores (possible range 0-100) were adjusted for protocols missing fewer than five items with the method recommended by Paulhus (1998).

^aUsing Bonferroni correction of α/n , or .05/3, *p* required for significance is .017.

^bCondition X Scale violated sphericity, so Greenhouse-Geisser correction is provided.

Table 14

Differences between Honest and Defensive Groups on PDS Scales

PDS Scale	Honest				Defensive				<i>d</i>
	<i>M</i>	<i>SD</i>	95% CI		<i>M</i>	<i>SD</i>	95% CI		
			LL	UL			LL	UL	
IM	50.30	11.35	47.14	53.46	68.42	18.71	63.21	73.63	1.17
SDE	59.73	8.47	57.37	62.09	68.64	8.08	66.39	70.89	1.08
SDD	58.13	9.99	55.35	60.91	71.93	13.96	68.04	75.82	1.14

Note. Continuous scores (possible range 0-100) were adjusted for protocols missing fewer than five items with the method recommended by Paulhus (1998).

Cut Scores for Defensive Responding

Empirically validated cut scores on standardized measures can provide reliable ways of determining response styles. Having established the ability of dually diagnosed inpatients to conceal Axis II symptomatology, the third research question explored whether this defensiveness could be identified with the existing measures in the current study. High sensitivity and positive predictive power (PPP) are the most important utility estimates to consider when identifying

response styles (Rogers, 2008b), but PPP is affected by the base rate of the response style in question. The current use of a within-subjects design created a base rate of 50%. Without knowing the actual base rate of defensiveness (see Discussion), additional estimates of 25% and 75% were implemented based on the range of base rates found in Baer and Miller's (2002) meta-analysis of faking good studies on the MMPI-2. This wide range also provides an indication of the strategies' utility in varying situations and allows for linear interpolation of utility estimates at other base rates.

Paulhus Deception Scales

The IM scale, measuring intentional response distortion, provides a method for identifying both faking good and faking bad. Based on the idea that validity scales can be used bidirectionally (Greene, 2010), the PDS (Paulhus, 1998) provides high (faking good) and low (faking bad) cut scores. When scored dichotomously, the IM scale has cut scores identifying protocols that “may be invalid” and that are “probably invalid,” resulting in four possible classifications (see Table 15). As with their continuous scores, described previously, the current sample evidenced a significant increase in IM dichotomous scores ($M = 11.06$, $SD = 5.48$), placing the defensive group mean in the “may be invalid” range for faking good, $t(51) = -7.34$, $p < .001$, $d = 1.29$.

Although malingering is not of interest in the current study, the availability of faking bad cut scores allowed for the examination of opposite misses. Of considerable concern, eight inpatients were misclassified as potentially faking bad when responding honestly or defensively. These misclassifications, especially the opposite misses, are difficult to explain; neither of the two defensive individuals evidenced increases in their personality disorder symptomatology under defensive responding, but both were classified as probably faking bad under honest

responding and diagnosed with Antisocial Personality Disorder (one also received a diagnosis of Borderline Personality Disorder). Thus, their classification as faking bad while responding defensively is likely due to individual differences in responding to the simulation instructions.

Table 15

Classification Rates under Honest and Defensive Conditions Based on IM Cut Scores

Condition	Fake Good ($N = 52$)		Fake Bad ($N = 52$)	
	May be invalid $9 \leq IM \leq 12$	Probably invalid $IM \geq 13$	May be invalid $IM = 1$	Probably invalid $IM = 0$
Honest	8	1	3	3
Defensive	16	18	2	0

The current study tested the applicability of PDS cut scores for defensive responding among inpatients. The majority of existing research on the PDS has used undergraduate samples, limiting its applicability to clinical populations. Additionally, the PDS only offers one set of cut scores, to be applied universally regardless of population. Fortunately, IM scores for dually diagnosed inpatients in the honest condition ($M = 5.27$, $SD = 3.17$) were not remarkably different than Paulhus' (1998) reported norms for college students ($M = 5.3$, $SD = 3.5$; $d = 0.01$), prisoners ($M = 5.3$, $SD = 3.6$; $d = 0.36$), military recruits ($M = 6.3$, $SD = 3.9$; $d = 0.28$), or the general population ($M = 6.7$, $SD = 4.0$; $d = 0.36$). Thus, the general cut scores are based on scores similar to those found in the current sample under honest conditions.

With two cut scores of varying stringency, the lower threshold acts as a broad net, resulting in greater sensitivity at the cost of misclassifying honest responders (i.e., lower PPP). Using the PDS' lower cut scores (see Tables 15 and 16), nearly two-thirds (65.4%) of the defensive responders were correctly identified. However, 28.8% of the honest responders were misclassified as either faking good (17.3%) or faking bad (11.5%). Much like the two-pronged approach to assessment, an initial screen over-identifies individuals for further examination.

Table 16

Utility of IM Faking Good Cut Scores Classifying Defensiveness among Dually Diagnosed Inpatients

Cut Scores	Sens.	Spec.	25% Base Rate			50% Base Rate			75% Base Rate		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
May be invalid (≥ 9)	.65	.83	.56	.88	.78	.79	.70	.74	.92	.44	.70
Probably invalid (≥ 13)	.35	.98	.86	.82	.82	.95	.60	.66	.98	.33	.51

Note. Sens. = sensitivity; Spec. = specificity; PPP = positive predictive power; NPP = negative predictive power; OCC = overall correct classification.

Predictably, the more stringent “probably invalid” cut score (see Table 16) was far less sensitive than the lower threshold, making it an appropriate second step in the two-pronged approach to identification. This sacrifice in sensitivity resulted in greater accuracy among those classified as defensive (i.e., PPP). Whereas the lower cut score misclassified 28.8% of honest inpatients as either faking good or faking bad, the higher cut score misclassified only 7.7% (see Table 15). Despite its accuracy, this stringent cut score is inefficient on its own due to its high false negative rate (65%, or $1 - \text{sensitivity}$). Among this dually diagnosed sample, a two-pronged approach of screening and refining classification resulted in the most efficient and accurate identification of defensive responders. However, using the PDS as a supplemental measure to detect defensive responding, while moderately successful, would add 10-15 minutes onto a clinical assessment—time that might be better spent gathering clinical information or attending to other issues.

Axis II Measures

Embedded detection strategies can streamline the assessment process in busy clinical settings, eliminating the need for a secondary measures and increasing the likelihood of implementation. Detection strategies were incorporated into the current study’s existing measures of Axis II symptomatology. As an initial step, two detection strategies were used to construct defensiveness scales based on the current sample’s Axis II prevalence rates. This

process provided the opportunity to match detection strategies to the specific response style of defensiveness. As noted by Rogers (2008b), pairing detection strategies with response styles is “essential” (p. 16) for interpreting results. Defensive responding, the intentional denial or minimization of psychiatric symptoms (Rogers, 2008a), is best identified with strategies that directly address this response style. Both strategies described here utilize Rogers’ (2008b) denial of patient characteristics strategy, though their specific methodology differs. Additionally, with the success of the two-pronged approach to identification on the PDS, a similar approach was undertaken in the development of these embedded strategies. Having two available cut scores allows clinicians to identify clients suspected of defensive responding with the lower threshold and make accurate decisions with the upper threshold, leaving those in the middle in an “indeterminate” group. This tiered approach to classification is also used for feigning determinations using the SIRS-2 (Rogers, Sewell, & Gillard, 2010).

Indiscriminate Denial of Symptoms (IDS). Indiscriminate denial of symptoms refers to a detection strategy for defensive responding that can best be conceptualized (Rogers, 2008c) as categorically denying all or nearly all psychological problems and psychopathology. It identifies as defensive those patients endorsing an unusually low number of symptoms. Looking at honest response patterns across both measures (see Table 10), inpatients endorsed an average of 25% ($SD = .10$) of the SIDP-IV symptoms and 39% ($SD = .14$) of the SCID-II-PQ items.

IDS relies on the total SIDP-IV and SCID-II-PQ scores in the establishment of cut scores. In establishing cut scores for new scales, researchers typically rely on utility estimates to optimize cut scores. However, this approach often capitalizes on error variance, thereby inflating the estimates. For the current thesis, initial cut scores were created using standard deviations for the honest group. As a screen for defensive responding, the IDS Screen identifies respondents

with scores one *SD* below the average total score. To refine classification, the IDS Indicator applies only to respondents with scores two *SDs* below these averages, or those in the lowest 2% of the population. For this strategy, protocols with fewer than 15% or 5% reported symptoms on the SIDP-IV and 25% or 11% on the SCID-II-PQ were flagged as may be and probably invalid (see Table 17).

As with the PDS cut scores, the IDS Screen provided excellent sensitivity on both the SIDP-IV and the SCID-II-PQ while the IDS Indicator provided enhanced PPP. Overall, this strategy was more successful on the SIDP-IV, but its success on the SCID-II-PQ suggests it can easily be incorporated during the screening process of Axis II assessments. This method has the added advantage of increased accessibility; many test manuals publish norms (including scale *Ms* and *SDs*) for various populations, allowing for the ad hoc creation of indiscriminate denial of symptoms scales on different measures.

Table 17

<i>Utility of IDS Strategy Classifying Defensiveness among Dually Diagnosed Inpatients</i>											
SIDP-IV	Sens.	Spec.	25% Base Rate			50% Base Rate			75% Base Rate		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
IDS Screen (< 15%)	.81	.90	.74	.93	.88	.89	.83	.86	.96	.61	.83
IDS Indicator (< 5%)	.25	1.00	1.00	.80	.81	1.00	.57	.63	1.00	.31	.44

SCID-II-PQ	Sens.	Spec.	25% Base Rate			50% Base Rate			75% Base Rate		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
IDS Screen (< 25%)	.64	.85	.58	.87	.79	.81	.70	.74	.93	.44	.69
IDS Indicator (< 11%)	.19	.98	.77	.79	.78	.91	.55	.59	.97	.29	.39

Note. Sens. = sensitivity; Spec. = specificity; PPP = positive predictive power; NPP = negative predictive power; OCC = overall correct classification; IDS = indiscriminate denial of symptoms.

Denial of Commonly Reported Symptoms (D-CRS). D-CRS (Rogers, 2008c) constitutes a second defensiveness strategy with greater sophistication than IDS. This strategy focuses specifically on the denial of common, everyday problems, which are likely to occur among both

clinical and community samples. Detection strategies for D-CRS focus on individual symptoms, unlike those for IDS that assess than overall response rates. This strategy was not implemented on the SCID-II-PQ; because it is intended as a screen, more than half (52.1%) the SCID-II-PQ items were commonly endorsed. Moreover, the likelihood of missing items on a self-report measure limits the applicability of this strategy if any of the D-CRS items were left unanswered.

The creation of a D-CRS scale on the SIDP-IV followed the same methodology used by Rogers (2008c) on an Axis I interview-based diagnostic measure, and cut scores were selected in the same method described for IDS (then rounded down to the nearest whole number). The SIDP-IV D-CRS scale is comprised of 13 *DSM-IV-TR* symptoms (see Table 18) that were present in 40% or more of the dually diagnosed inpatients. Incorporating both reported and observed symptomatology, honest responders averaged 7.19 ($SD = 2.64$) of these symptoms, whereas defensive responders averaged just 2.65 ($SD = 2.66$).

Consistent with its increased sophistication, the D-CRS strategy (see Table 19) generally outperformed IDS with better sensitivity and accuracy. While PPP was slightly lower for D-CRS, NPP increased; in other words, the D-CRS Indicator was accurate in classification of both honest and defensive responders and had greater OCC rates (56-84% for D-CRS versus 44-81% for IDS). Of the three methods developed in the current thesis, D-CRS was most effective in identifying potential defensive responders and accurately classifying all respondents.

Table 18

Commonly Reported Personality Disorder Symptoms Among Dually Diagnosed Inpatients

Diagnosis	DSM-IV-TR Symptom as Described on the SIDP-IV (Pfohl et al., 1997)	%
Borderline	Impulsivity in at least two areas that are potentially self-damaging	86.5
Antisocial	Repeatedly performing acts that are grounds for arrest	76.9
Antisocial	Consistent irresponsibility	63.5
Antisocial	Impulsivity or failure to plan ahead	59.6
Obs-Comp	Is reluctant to delegate tasks or work to others	59.6
Antisocial	Evidence of Conduct Disorder before age 15 years	55.8
Borderline	Affective instability due to a marked reactivity of mood	51.9
Borderline	Inappropriate, intense anger or difficulty controlling anger	50.0
Obs-Comp	Shows rigidity and stubbornness	46.2
Borderline	A pattern of unstable and intense interpersonal relationships	44.2
Borderline	Chronic feelings of emptiness	42.3
Borderline	Recurrent suicidal behavior, gestures, or threat, or self-mutilating behavior	42.3
Histrionic	Interaction with others is often characterized by inappropriate sexually seductive or provocative behavior	40.4

Note. Obs-Comp = Obsessive-Compulsive. Symptoms that were rated as present or strongly present in at least 40% of the current sample.

Table 19

Utility of SIDP-IV D-CRS Scale Classifying Defensiveness among Dually Diagnosed Inpatients

Cut Scores	Sens.	Spec.	25% Base Rate			50% Base Rate			75% Base Rate		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
D-CRS Screen (≤ 4)	.85	.85	.65	.94	.85	.85	.85	.85	.94	.65	.85
D-CRS Indicator (≤ 1)	.42	.98	.88	.84	.84	.96	.63	.70	.96	.36	.56

Note. Sens. = sensitivity; Spec. = specificity; PPP = positive predictive power; NPP = negative predictive power; OCC = overall correct classification; D-CRS = denial of commonly reported symptoms.

The third research question provided several fruitful areas for further research on the detection of defensive responding. Initial analyses, based on the current sample of dually diagnosed inpatients, established two built-in strategies for identifying defensive responding on Axis II measures. On both the SCID-II-PQ and the SIDP-IV, a two-pronged approach for implementing these strategies proved moderately successful in identifying and classifying defensive responding. It is important to note that the exclusion criteria (i.e., < 10 SIDP-IV symptoms) necessarily affected the creation and utility of detection strategies by artificially inflating average honest scores. This will be addressed in more detail in the Limitations section

of the Discussion.

Narcissism and Self-Deceptive Enhancement

Narcissism and self-deceptive enhancement share similar characteristics, namely a lack of insight and overly favorable self-presentation, suggesting that the two clinical constructs might be strongly correlated (Pauls & Crost, 2004; Salekin, 2000). In exploring this supplementary research question, the current sample of dually diagnosed inpatients exhibited a small but significant correlation ($r = .32, p = .02$) between honest SDE scores and dimensional narcissism scores on the SIDP-IV. However, it is important to observe that narcissism was uncommon in the current sample (range for SIDP-IV narcissism dimensional score was 0 – 15). In fact, 28.8% had scores of three or less (i.e., less than 2 symptoms) for Narcissistic Personality Disorder. Therefore, its restricted range could be contributing to the low correlation. This supposition was tested by excluding those 15 individuals without significant narcissistic symptoms. Unexpectedly, the correlation is even smaller ($r = .14, p = .42$). Initial exploration of this question, though limited by minimal representation of narcissism in the current sample, does not appear to support the claim that narcissism and SDE are overlapping constructs. Alternatively, this low correlation and lack of convergent validity may be explained by poor construct validity (Pauls & Crost, 2004) of the SDE scale.

Self-Reported Ability to Portray a Favorable Impression

Insight into deceptive abilities has been previously examined for certain response styles but not specifically for the suppression of Axis II traits. Most inpatients believed they were successful in achieving placement onto the more desirable unit (the goal described in the Phase II scenario) and had put forth good or excellent effort during the study. Perceived success could not be effectively evaluated because all but six inpatients reported high perceived success,

eliminating substantial variance in the independent variable. Reported effort was also limited by its restricted range (21 participants³ reported medium or good effort, and 31 reported excellent effort). Those reporting lower effort did not evidence significantly smaller dimensional score reductions on the SIDP-IV than those who reported excellent effort, $t(49) = -.287, p = .775$. Further supporting this finding, nine of the 11 participants who failed to uniformly decrease their personality disorder scores reported excellent effort (with the other two reporting good effort). Thus, in response to the second supplementary research question, perceived effort does not appear to predict actual success in defensive responding.

³ One participant reported medium effort. This individual was combined with the good effort group via contrast testing.

CHAPTER 4

DISCUSSION

Clinical psychologists are confronted with multiple challenges in the assessment of Axis I and Axis II diagnoses. As noted in the introduction, respondents are often less than forthcoming in their presentation of symptoms and psychological characteristics. Given the stigmatization of Axis II features and the compounding effects of narcissism or poor insight, the exact breadth of this problem in personality disorder evaluations is immeasurable. Of particular concern are issues of defensiveness and other related response styles for creating a favorable self-presentation. Although studied extensively with general personality and psychopathology measures, defensiveness has not been subjected to systematic research as it applies to diagnostic assessments. With estimates (Verheul et al., 2000; Zimmerman et al., 2005) that approximately half of the clinical population has at least one personality disorder, understanding the effects of defensive responding on Axis II assessments is essential.

Work on defensive responding in personality functioning has not been extended to formal Axis I and Axis II measures. For example, a PsycINFO search of “personality” and relevant search terms⁴ revealed 785 studies looking at styles of favorable self-presentation, including defensiveness. In stark contrast, a similar search for “diagnostic” and the same response style terms yielded just 37 studies. This issue of research’s general neglect of defensiveness in personality disorders will be discussed.

Multiscale inventories provide us with an initial, if imperfect, overview of the prevalence of defensiveness on Axis II measures. Among two groups ($n_s = 80$) of adult outpatients

⁴ Search included any of the following terms: defensiveness, favorable self-presentation, or faking good. To avoid counting non-empirical publications, search was limited to English-language peer-reviewed journals.

responding under standard instructions, means on the MCMI's Scale Y, or the Desirability Index, achieved base rates (BR) of 51.8 ($SD = 22.5$) and 52.6 ($SD = 22.4$; Daubert & Metzler, 2000). With the MCMI-III base rates set upon normative data (Millon et al., 1997), these figures indicate the outpatients performed substantially below the established median of BR 60. In contrast, outpatients instructed to "fake good" were substantially elevated with an average of 73.5 ($SD = 22.8$) and those instructed to "fake bad" had a very low average of 33.0 ($SD = 22.4$). While providing an estimate of defensive responding among clinical outpatients is relevant, this example does not assess the influence of real-life motivation to provide overly favorable self-presentations. Reviewing five MMPI-2 studies with child custody litigants and personnel samples, Baer and Miller (2002, p. 22) established .30 (range .20 – .74) as the "best available estimate" for the prevalence of underreporting. In child custody evaluations ($N = 259$), a situation wherein parents are highly motivated to appear well-adjusted, McCann and colleagues (2001) found Scale Y averages in the clinical range ($M = 75.56$, $SD = 13.06$), placing them in the upper quartile among the normative sample (Millon et al., 1997). Clearly, situational effects play a profound role in the likelihood of defensive responding.

Beyond prevalence, psychologists must take into account the potential impact of undisclosed personality pathology on the accurate diagnosis and effective treatment of clinical populations. As highlighted in the Introduction, clinicians generally have failed to use standardized assessment methodology and therefore have missed many Axis II diagnoses (Westen, 1997). On this issue, Tyrer and Mulder (2006) raised the question of whether undiagnosed personality disorders adversely impact clinical outcomes. Supporting this clinical hypothesis, an early review of psychotherapy outcomes determined that "general personality functioning" is directly related to treatment outcome such that "the sicker he is to begin with, the

poorer the outcome” (Luborsky, Auerbach, Chandler, Cohen, & Bachrach, 1971, pp. 147-148). With these studies taking place prior to the creation of formal personality disorder criteria, “general personality functioning” included a broad range of variables.

Axis II disorders can lead to increased service utilization in addition to less promising treatment outcomes. For instance, Bender et al. (2001) found that patients with personality disorders utilized significantly more services than those with major depression. From a motivational perspective, researchers (Bender et al., 2001; Van Beek & Verheul, 2008) have suggested that Axis II disorders contribute to treatment-seeking above and beyond Axis I disorders. More focused studies have followed, with research moving toward the interactions of specific disorders, such as (a) Antisocial Personality Disorder and alcohol use disorders and (b) Avoidant Personality Disorder and bulimia nervosa. In a summary of such studies, Tyrer and Mulder (2006) found generally small, adverse effects on the treatment of Axis I disorders.

This chapter begins with a historical perspective of the development of personality measures as they relate to response styles. The first section outlines the tension between direct and indirect approaches to clinical assessment, spanning the early work with projective measures to more modern, standardized methods. The second section builds on the development of Axis II diagnostic measures but with a specific focus on the assessment of response styles via specialized scales and other indicators. In light of past research, the third section considers the current research findings in relationship to Axis II diagnosis. It critically examines the susceptibility of both a self-report and an interview to defensive responding. Given the vulnerability of Axis II measures, this section also explores the implementation of detection strategies. The chapter concludes with a discussion of this study’s limitations and implications, including methodological improvements in test development to diminish susceptibility to

response distortion, taking into consideration the upcoming diagnostic shift (APA, 2011) from categorical to dimensional assessment.

The Role of Transparency and Structure in Personality Assessment

In the evolution of personality assessment, two core qualities aptly describe most forms of personality measurement: transparency and structure. Ranging from indirect and unstructured projective measures to overly direct and detail-orientated questionnaires, transparency and structure have historically been inversely related. As with all psychological measures, personality researchers have strived to find test formats that balance optimum performance with ease of use. While enhancing structure has been largely positive for reliability and validity, increased transparency also provides increased opportunity for socially desirable responding. Structured and semi-structured interviews straddle the divide and are positioned toward the middle of both continua. Consideration of transparency and structure has played a key role in the evaluation of distorted responding; many investigators assume that transparent and structured measures are more susceptible to distortion than less transparent, less structured measures (Butcher, 2010; Meehl, 1945; Meehl & Hathaway, 1946)

Psychologists at the beginning of the twentieth century developed highly innovative approaches to assess personality. Influenced by the popularity of psychoanalysis, projective techniques were developed to measure core aspects of personality structure (Lilienfeld, Wood, & Garb, 2000). Although originally used in the classical Freudian sense, the use of the term “projective” was broadened to simply mean that responses to ambiguous stimuli are influenced by personal characteristics, allowing test administrators insights about respondents’ psychological processes (Exner, 2003; Meyer & Kurtz, 2006). Designed to tap into constructs seen as largely beyond self awareness, projective assessment methods were traditionally

considered highly resistant to response distortion (Meehl, 1945). Moreover, projective methods did not rely primarily on respondents' insight or willingness to disclose. In contrast to these traditional formulations, Sewell's empirical review (2008) provided several salient examples wherein respondents were able to significantly alter their presentation on projective measures; such intentional distortions may limit the usefulness and interpretability of projective measures with non-cooperating clients.

Amid the arising controversy about projective measures' utility (see Lilienfeld et al., 2000, for an extensive review) and the movement toward increased empiricism in psychology, objective testing gained considerable traction in the 1930s and 40s. Early objective measures were constructed on a rational basis with the inclusion of items that, *prima facie*, captured the intended construct and made it readily identifiable to the test-taker (Butcher, 2010; Greene, 2000; Holden & Jackson, 1979). In contrast to projective measures, "structured personality tests" (Meehl, 1945) asked specific questions about behaviors and characteristics. However, objective questions do not necessarily yield objective responses. Representing a strikingly negative view of respondents' ability to answer honestly, Stagner (1934) asserted the following: "In a few cases I attempted to get self-ratings from the students, but such complete mental confusion appeared that this attempt was abandoned. The self-rating technique is better suited to more sophisticated individuals." (p. 414). Honest and accurate answers, as Meehl observed, require both insight into and willingness to fully disclose the characteristic or symptom in question. Thus, the introduction of rationally constructed scales and face valid or transparent items provided respondents with the choice to respond either honestly or in a distorted manner. To ameliorate this problem, empirical test construction (Holden & Jackson, 1979; Meehl, 1945; Meehl & Hathaway, 1946) eschewed face validity in favor of subtle item content that reliably

distinguished between criterion groups.

Empirical test construction, commonly known as empirical keying, provided objective tests with limited transparency that was previously available only with projective measures. At its extreme, items could be selected on a purely statistical basis, making item content less important than item performance (Holden & Jackson, 1979). Hathaway and McKinley (1940) were among the first to use empirical keying when they developed the Minnesota Multiphasic Personality Inventory (MMPI). Regardless of item content, items that discriminated between respondents with and without a particular personality constellation were used to create a scale. As scale items are not necessarily homogenous or face valid, empirical keying may obscure the identity of the construct being measured.

With the rise of clinical psychology and the burgeoning role of the *DSM*, psychometric measures became increasingly popular as a key component in diagnostic decision-making. Dedicated Axis I diagnostic measures tend to be highly structured and transparent, much like the early objective measures (Westen, 1997). Similarly, Axis II measures map onto formal DSM diagnostic inclusion criteria, starting with DSM-III (APA, 1980) and continuing to the most recent edition, DSM-IV-TR (APA, 2000). Westen contends that transparency may be useful for assessing Axis I disorders, but using direct questions in the assessment of personality disorders is, at the very least, not ideal due to the considerable insight required for honest and objective self-appraisals. As evidenced in the current study, direct measures of Axis II symptomatology appear to be highly susceptible to intentional response distortion.

Axis II assessments often adopt a multistep process to avoid overreliance on clients' insight and objectivity. For example, Widiger and Samuel (2005) recommended administering a self-report inventory to simply screen for clinical levels of pathology but then confirm Axis II

diagnoses with a clinician-administered structured interview. Structured diagnostic interviews often parallel the diagnostic inventories inasmuch as both ask directly about criteria; however, interviews provide the opportunity for unstructured responses, including salient clinical examples, rather than completely relying on the face value of responses (Widiger & Axelrod, 1995). They may also reduce the likelihood that a test-taker will misunderstand a question because a skilled clinician may detect confusion or serious misunderstandings. To further minimize the effects of possibly distorted self-appraisals, semistructured interviews require skilled practitioners to apply operationally-defined ratings in determining when respondents meet DSM criteria.

While multimethod assessments provide the chance for convergent validity, they do not necessarily reduce reliance on respondents' honest and objective self-appraisals. For the purpose of reducing symptom denial or minimization based on social desirability, Pfohl and colleagues (1997) carefully crafted SIDP-IV questions in a way that makes them sound less pathological. For example, asking, "What kinds of things do you enjoy?" was used to assess the Schizoid criterion of "takes pleasure in few, if any, activities" (p. 1). In contrast to the SIDP-IV, the Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sickel, & Young, 1996) consists mostly of yes/no questions focused directly on pathology. For the same Schizoid criterion, it queries, "Have you found that most of your activities don't give you much pleasure?"

Personality measures can also be designed to minimize the impact of social desirability by asking respondents how others view them (i.e., metaperceptions). As Oltmanns, Gleason, Klonsky, and Turkheimer (2005) demonstrated, however, the value of such metaperceptions for personality functioning may be limited: individuals tend to think that others share their views.

Inaccurate metaperceptions likely undermine the clinical utility of this assessment method. The twin problems of inaccuracy and inconsistency can be reduced by obtaining collateral information from others (e.g., First et al., 1997; Widiger & Samuel, 2005). Many diagnostic measures have been or could be adapted for use with informants (Klonsky, Oltmanns, & Turkheimer, 2002; Widiger & Samuel, 2005), but some strictly informant-based measures (e.g., Standardized Assessment of Personality; Mann, Jenkins, Cutting, & Cowen, 1981) have also been developed.

The evolution of Axis II measures has brought about several clinical improvements that may affect the accuracy of personality diagnoses. Despite these advances, researchers (Westen, 1997; Zimmerman & Mattia, 1999; Zimmerman et al., 2005) have observed that many clinicians do not rely on diagnostic measures, direct questions, or informant reports to assess Axis II disorders, even if they use similar approaches for Axis I diagnoses. Rather, they prefer indirect methods, such as observing clients' behavior and listening to their anecdotes (Westen, 1997), that are prone to missed diagnoses (Zimmerman & Mattia, 1999). As an example, only 2.8% (2 of 71) of the current study's participants had personality disorder diagnoses in their hospital charts, but 57.7% (41 of 71) met diagnostic criteria for at least one personality disorder using the SIDP-IV.⁵ Amidst the practical constraints of clinical settings, the accuracy of diagnostic testing is often weighed against its the efficiency of unstructured methods. Thus, the problem becomes one of developing methods that are both efficient and accurate, methods that clinicians are likely to use regularly. A first step toward solution is finding methods to detect intentional response distortion, a glaring vulnerability of current diagnostic measures.

⁵ Interestingly, one of the two diagnoses was not supported by SIDP-IV results (the participant did not meet diagnostic criteria for Borderline Personality Disorder, as indicated in the chart).

Favorable Self-Presentation in Personality Assessment

Since psychopathology's beginnings, maladaptive personality traits have been the subject of much stigma. Freud and his contemporaries viewed personality pathology as “weakness of character or willfully offensive or socially deviant behavior” (Oldham, 2005, p. 373). With their inclusion as seemingly permanent personality “disturbances” and “defects” in the first DSM (APA, 1952, p. 34), personality pathology has a long history of stigmatizing language (Oldham, 2005). Despite attempts to shed stigmatizing terminology and negative personal attributions, the fact remains that personality pathology is comprised of maladaptive behaviors that cause interpersonal problems. In effect, this stigmatization leads individuals, including service providers, to make negative personal attributions about those with personality disorders (Lewis & Appleby, 1988).

The implicit assumptions about morality and responsibility in personality disorders further contribute to their stigmatization. On this point, Charland (2006) contends that Antisocial, Borderline, Histrionic, and Narcissistic Personality Disorders, as conceptualized in the DSM-IV, are imbued with moral judgments. The “morally loaded” (Charland, 2006, p. 119) criteria presuppose that an individual diagnosed with a Cluster B disorder has engaged in immoral activities, such as lying or manipulating others. Whether individuals with personality disorders are responsible for their actions has become the focus of philosophical (Zachar & Potter, 2010) and political (Mohino, Pujol, & Idiaquez, 2011) debate. With more medicalized disorders (i.e., disorders with clear etiologies and treatments), if the actions are viewed as a product of the disorder, the person appears to be less at fault or even blameless. However, personality disorders are distinct from Axis I disorders or medical conditions in their pathologizing of variations on normal functioning (Widiger, 1997). In line with Charland's

argument, such appreciable deviations from normality result in increased attribution of blame toward the person, not the disorder (Aviram et al., 2006). Antisocial Personality Disorder (APD) provides a compelling example of this bias; in the American Law Institute's (1962) standard for insanity, APD and chronic criminal behavior are specifically excluded from the criteria for mental diseases or defects as the basis for exculpation. The child custody literature (McCann et al., 2001) provides a further example of how concealment of these morally loaded behaviors is very likely when individuals are being judged or evaluated.

Given the high levels of stigmatization, it is likely that many examinees with personality pathology endeavor to “protect certain aspects of the self (e.g., psychological impairment) from being revealed to others” (Sewell, 2008, p. 215) during psychological assessment. Despite this observation, the examination of defensive responding on personality disorder measures has been largely absent from the literature, except for focused research on multiscale inventories like the MCMI and MMPI. In my opinion, the motivation for defensiveness combined with the vulnerability of face valid questions to distorted responding would suggest that Axis II assessments would be substantially strengthened by the inclusion of well-validated methods for detecting defensiveness. However, validity scales are not common on diagnostic measures for personality disorders (see Table 20), and defensive responding could easily go undetected.

Table 20

Methods for Detecting Defensiveness on Axis II Diagnostic Measures

Scale	Source(s)	Defensiveness	Validated with Sample	
			Community	Clinical
Structured Interviews				
AUDADIS-IV	Client	No	--	--
DIPD	Client	No	--	--
PDE	Client	No	--	--
PDI-IV	Client	No	--	--
SIDP-IV	Client	No ^a	--	-- ^a
SAP	Informant	No	--	--
PAS	Client & Informant	No	--	--
Structured Interviews Supplemented with Screens				
IPDE & IPDE-SQ	Client	No	--	--
IPDS ^b	Client	No	--	--
SAPAS or SAPAS-SR ^c	Client	No	--	--
SCID-II & SCID-II-PQ	Client	No ^a	--	-- ^a
Inventories and Questionnaires				
WISPI-IV	Client	No	--	--
OMNI-IV	Client	No	--	--
PDQ-4	Client	Yes	Yes	No
MCMI-III	Client	Yes	Yes	Yes
SNAP-2 ^d	Client	Yes	Yes	No
MAPP ^e	Client & Informant	No	--	--
CATI	Client & Informant	Yes	No	No
Rating Scales				
SWAP-200	Professional	No	--	--
PAF	Professional	No	--	--

Note. AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (Grant, Dawson, & Hasin, 2001); DIPD = Diagnostic Interview for DSM Personality Disorders (Zanarini, Frankenburg, Sickel, & Young, 1996); PDE = Personality Disorder Examination (Loranger, 1988); PDI-IV = Personality Disorder Interview-IV (Widiger, Mangine, Corbitt, Ellis, & Thomas, 1995); SIDP-IV = Structured Interview for DSM-IV Personality (Pfohl, Blum, & Zimmerman, 1997); SAP = Standardized Assessment of Personality (Mann, Jenkins, Cutting, & Cowen, 1981); PAS = Personality Assessment Schedule (Tyler, Alexander, & Ferguson, 1988); IPDE = International Personality Disorder Examination (Loranger, 1999); IPDE-SQ = International Personality Disorder Examination – Screening Questionnaire (Loranger, 1999); IPDS = Iowa Personality Disorder Screen (Langbehn et al., 1999); SAPAS = Standardized Assessment of Personality – Abbreviated Scale (Moran et al., 2003); SAPAS-SR = Self-report Standardized Assessment of Personality – Abbreviated Scale (Germans, Van Heck, Moran, & Hodiarnont, 2008); SCID-II = Structured Clinical Interview for DSM-IV Axis II Disorders (First, Gibbon, Spitzer, Williams, & Benjamin, 1997); SCID-II-PQ = Structured Clinical Interview for DSM-IV Axis II Disorders – Personality Questionnaire (First et al., 1997); WISPI-IV = Wisconsin Personality Disorders Inventory-IV (Klein & Benjamin, 1996); OMNI-IV = OMNI-IV Personality Disorder Inventory (Loranger, 2002); PDQ-4 = Personality Diagnostic Questionnaire-4 (Hyler, 1994); MCMI-III = Millon Clinical Multiaxial Inventory-III (Millon, Davis, & Millon, 1997); SNAP-2 = Schedule for Nonadaptive and Adaptive Personality-2 (Clark, Simms, Wu, & Casillas, in press); MAPP = Multi-Source Assessment

of Personality Pathology (Oltmanns, Gleason, Klonsky, & Turkheimer, 2006); CATI = Coolidge Axis II Inventory (Coolidge & Merwin, 1992); SWAP-200 = Shedler-Westen Assessment Procedure-200; PAF = Personality Assessment Form (Pilkonis & Frank, 1988).

^aMethods for detecting defensiveness on the SIDP-IV and SCID-II-PQ were developed on a clinical sample as part of the current study.

^bThe IPDS is a brief interview, derived from questions from an earlier version of the SIDP-IV; it has not been updated.

^cThe SAPAS, a brief client interview, and SAPAS-SR, a brief self-report, supplement the SAP. The measures were listed separately to indicate the different use of sources

^dValidity scales on the SNAP-2's precursor, the SNAP, have been subjected to simulation design.

^eThe MAPP is also available in "peer nomination" and "expected peer" forms.

Structured Axis II interviews, while occasionally capitalizing on multiple sources of data, have routinely relied on client responses. In line with assessment recommendations (Widiger & Samuel, 2005), a few interviews have incorporated screening questionnaires to be used prior to the interview (see Table 20). As noted by Widiger and Samuel (2005) and demonstrated in Table 20, Axis II interviews rely solely on clinical judgment to detect aberrant response styles, perhaps with the tacit assumption (Widiger & Axelrod, 1995) that interviews for personality disorders cannot be faked or that clinicians will have the ability to detect dissimulation. The current study tested the first premise, finding that Axis II interviews are very susceptible to response distortion. Regarding clinician detection, Buller, Strzyzewski, and Hunsaker (1991) demonstrated that individuals engaged in conversation tend to believe the speaker more than observers do. Serving as an analogue for administering a psychological assessment, this finding suggests that clinicians directly involved in the assessment may have a diminished ability to accurately detect response distortion. Pending replication and extension to different interviews and clinical settings, initial evidence appears to undermine these tacit assumptions and support the need to establish validity scales on Axis II interviews.

Assessments based on multiple sources reduce dependence on examinees' honesty and insight and, potentially, increase validity. Some measures (see Table 20) formally incorporate

these sources with informant-specific forms and instructions (e.g., SAP, PAS, MAPP, and CATI). Even if not indicated, however, most measures can be adapted for use with informants (Klonsky et al., 2002; Widiger & Samuel, 2005). Supporting the use of dual perspectives, Klein (2003) argued that both self and other sources should be used when making diagnoses because each provides unique information and incremental validity in predicting treatment success.

The most interesting finding in Table 20 involves the formal assessment of defensive responding via validity scales. For the category of Inventories and Questionnaires, four of the seven measures have developed scales for defensiveness. Remarkably, none of the interviews, screens, or rating scales include formal detection strategies. The IDS and D-CRS strategies developed in the current thesis mark the first use of validity scales on Axis II diagnostic interviews and screens. The effectiveness of defensiveness scales with clinical and community samples is described in the next section.

Defensive Responding on Personality Disorder Assessments

The assessment of response styles typically occurs after the first generation of diagnostic measures and is a multistep process (Rogers, 2008d). Compared with the assessment of major mental disorders, the development of diagnostic measures for Axis II disorders is a relatively young field. However, more than 60 years after Meehl and Hathaway's (1946) call to action for psychologists to "correct" and "overcome" (p. 526) personality measures' susceptibility to response distortion, the call is largely unanswered. The current thesis aimed to address this issue by testing the susceptibility of Axis II diagnostic measures and detecting defensive responding.

Susceptibility of Axis II Measures to Defensive Responding

An important question is whether examinees with substantial Axis II pathology can systematically distort their responses so as to appear well-adjusted. Building on existing research

(Bagby & Pajouhandeh, 1997; Fals-Stewart, 1995; Retzlaff et al., 1991), current results extended the simulation design to include two forms of Axis II diagnostic measures with a clinical sample. As expected, dually diagnosed inpatients were able to significantly decrease total psychopathology on both the SCID-II-PQ and the SIDP-IV ($d_s > 1$, see Tables 10 and 11). Boding well for its use as a screening instrument, the SCID-II-PQ (total $d = 1.06$; $M d = 0.73$) was slightly less susceptible than the SIDP-IV ($d = 1.40$; $M d = 0.86$). The effects of defensive responding on Axis II diagnoses was profound, with approximately 3 out of 4 (73.5%) previously diagnosed inpatients masking enough symptoms to evade diagnosis.

Systematic comparisons between past (Bagby & Pajouhandeh, 1997; Fals-Stewart, 1995; Retzlaff et al., 1991) and current studies of defensive responding on Axis II pathology require a common metric, specifically the use of effect sizes. Comparing across the four studies provides an initial evaluation of the patterns produced by defensive responding. All four measures evidenced several moderate ($d \geq 0.75$), large ($d \geq 1.25$), or very large ($d \geq 1.50$) effect sizes (see Table 21; Rogers, 2008b). In looking at diagnostic scales on each measure, average effect sizes ranged from 0.59 to 0.86 (grand $M d = 0.72$), with the SIDP-IV proving more susceptible (i.e., largest or second largest effect size) on seven of ten personality disorders and two of three clusters (see Table 21). This counterintuitive finding is in contrast to long-standing views that open-ended items are less susceptible to distortion (Meehl, 1945). Given the SIDP-IV's attempts to minimize the effects of social desirability (Pfohl et al., 1997), it is possible that other diagnostic interviews are just as susceptible, if not more so.

Table 21

Effects of Defensive Responding on Four Axis II Measures and Their Validity Scales

Scale	PDQ-4		MCMI-II		SCID-II-PQ		SIDP-IV		M
	Bagby & Pajouhandeh ^a	Retzlaff et al. ^b	Fals-Stewart ^c	Current Study ^d	Current Study ^e	Absolute Value			
Validity Scale	-0.71	-1.67	-0.57	--	-1.71	1.17			
Cluster A	0.38	0.71	0.56	0.89	0.89	0.69			
Paranoid	-0.47	-0.29	-1.06	-1.17	-0.94	0.79			
Schizoid	-0.25	-0.70	-0.36	-0.70	-0.64	0.53			
Schizotypal	-0.43	-1.14	-0.27	-0.80	-1.08	0.74			
Cluster B	0.56	0.68	0.96	0.71	0.89	0.79			
Antisocial	-0.01	-0.85	-1.24	-0.60	-1.38	0.82			
Borderline	-1.35	-1.63	-1.00	-1.48	-1.53	1.40			
Histrionic	-0.47	-0.25	-0.74	-0.28	-0.51	0.45			
Narcissistic	-0.40	0.00	-0.84	-0.46	-0.86	0.51			
Cluster C	0.82	0.77	0.55	0.61	0.55	0.66			
Avoidant	-1.01	-1.26	-0.63	-0.83	-0.74	0.89			
Dependent	-1.04	0.12	0.10	-0.31	-0.31	0.38			
Obsessive-Compulsive	-0.42	0.92	0.93	-0.70	-0.59	0.71			
M Axis II Diagnostic Scales (Absolute Value)	0.59	0.72	0.72	0.73	0.86	0.72			

Note. Cohen's *d* values were computed so negative values indicate defensive score decreases and positive values indicate defensive score increases as compared to honest scores. Student samples are shaded, clinical samples are unshaded.

^aBagby & Pajouhandeh (1997). This study used a between-subjects design, with samples consisting of 48 honest and 36 defensive university students (some protocols included missing items); Too Good scale used for validity scale.

^bRetzlaff, Sheehan, & Fiel (1991). This study used a between-subjects design, with samples consisting of 50 university students each in honest and defensive conditions; *d* computed using raw scores; Scale Y used for validity scale.

^cFals-Stewart (1995). This study used a between-subjects design, with samples consisting of 62 substance-users each in honest and defensive conditions; *d* computed using base rates; Scale Y used for validity scale.

^dSCID-II-PQ total yes responses used (some protocols included missing items). No validity scale included. Antisocial scale only assesses for Conduct Disorder, criterion C of Antisocial Personality Disorder.

^eSIDP-IV dimensional sums used. D-CRS scale used for validity scale.

Clinical and community samples are likely to approach personality disorder measures very differently. Starting with higher prevalence rates (Grant et al., 2004; Lenzenwenger et al., 2007; Verheul et al., 2000; Zimmerman et al., 2005), clinical samples have the opportunity for larger effect sizes. Among these four studies, Bagby and Pajouhandeh's (1997) and Retzlaff and

colleagues' (1991) used university students, while Fals-Stewart (1995) and the current study used individuals in treatment for substance use (participants in the current sample were dually diagnosed with a substance use disorder and a major mental disorder). Variance on clinical scales was lowest on the PDQ-4, but without comparison to a clinical sample it is difficult to discern whether this is due to the measure itself or the sample with which it was used. Interestingly, both the clinical and community studies using the MCMI-II produced similar average reductions, though the pattern differed by diagnosis.

Paradoxical effects were found for the two MCMI-II studies. On Dependent and Obsessive-Compulsive Personality Disorders, both MCMI-II samples evidenced increases under defensive responding. These findings were especially strong for the latter disorder. Given that all other scales showed decreases under defensive responding and that these increases occurred in both clinical and community samples, these MCMI-II scales might present an anomalous approach to personality disorder measurement. Alternatively, respondents might view these disorders as less socially unacceptable. The latter explanation is consistent with the other studies showing relatively low effect sizes for these disorders.

Disorders with the greatest stigmatization ought to demonstrate the largest change if respondents alter their answers based on the inherent social desirability of items (Edwards, 1953; Jackson, 1996; Pfohl et al., 1997). As Aviram and colleagues (2006) discussed, Borderline Personality Disorder is likely to be highly stigmatized due to its specific constellation of symptoms, including extreme volatility in interpersonal situations. Across the four studies in Table 21, Borderline Personality Disorder was the only diagnosis with an average effect size ($d = 1.40$) in the large range. While clinical samples (Fals-Stewart, 1995 and the current study) were elevated on Borderline scales under honest conditions, the university students (Bagby &

Pajouhandeh, 1997; Retzlaff et al., 1991) were not. Thus, these large and very large effect sizes are more likely due to the stigmatizing nature of Borderline Personality Disorder than due to its high prevalence in honest conditions. Other disorder-level findings suggest that clinical and community samples take different approaches to personality disorder measurement and possibly have different perceptions of what is acceptable or desirable. For Paranoid and Narcissistic Personality Disorders, clinical samples showed higher reductions, whereas student samples achieved larger reductions on Avoidant Personality Disorder and Cluster C in general. Assessing the desirability of each item during scale development (Edwards, 1953; Jackson, 1996) can help to understand this type of pattern in varying populations and assist in the development of effective validity scales.

Detection of Defensive Responding

This section begins with an overview of favorable self-presentation, which provides a conceptual basis for the assessment of defensive responding. Taxonomy of favorable self-presentation has long suffered from vague or overlapping definitions and indiscriminate classification. Most scales measuring favorable self-presentation fit into a two-axis model of social desirability, distinguishing between (a) intentional and unintentional response distortion and (b) enhancement of virtues and denial of deficits (see Figure 2). As described in Table 1, some response styles straddle these dimensions. For example, impression management, as described by Paulhus (1984, 1998), could include intentional denial of deficits, unintentional enhancement of virtues, or both. However, this model is useful in creating detection strategies to match specific response styles.

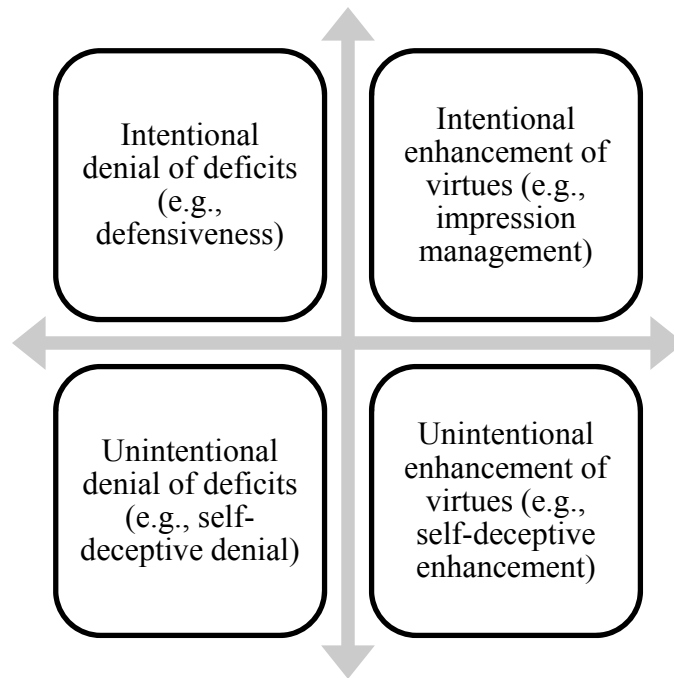


Figure 2. Two-axis model of favorable self-presentation, with intentionality on the y axis and type of distortion on the x axis.

The distinction between unintentional and intentional detection strategies began with Meehl and Hathaway’s (1946) development of the L scale for intentional distortion and the K correction for unintentional distortion, noting structured tests’ “even greater susceptibility to unconscious self-deception” (p. 525). These scales are also distinct in their type of content, with the L scale being far more transparent than the K correction as a function of their differing methods of development (Meehl & Hathaway, 1946). Lanyon (2001, 2004) and Paulhus (1984, 1998) both included self-deceptive response styles in their models of favorable self-presentation. In my opinion, however, unintentional self-deception is better conceptualized as lack of insight, consistent with Axis I disorders (e.g., psychotic disorders, Obsessive-Compulsive Disorder). In personality disorders, lack of insight would be classified as an enduring patterns related to cognition, such as perceptions of self and others (APA, 2004). Consistent with this approach,

Uziel (2010) argued that only intentional distortion counts as response bias in psychological testing. Unfortunately, continued stigmatization of personality pathology contributes to the attribution of blame toward individuals rather than their disorders (Aviram et al., 2006), and perpetuates the classification of unintentional self-deception as a response style rather than a characteristic.

Detection strategies and intentionality. The PDS is presented as a measure of intentional and unintentional response styles, but it appears better suited to measure more nuanced styles of intentional favorable responding. Simulation studies, instructing participants to *intentionally* distort their responses, provide an opportunity to assess the criterion validity of both intentional (IM) and unintentional (SDE and SDD) favorable self-presentation scales by comparing their responsiveness. In the current study, the two PDS scales designed to measure unintentional distortion (SDE and SDD) increased as much as the scale measuring intentional distortion (IM; all d s > 1, see Table 14). Pauls and Crost (2004) also found large, consistent increases for both SDE ($d = 1.51$) and IM ($d = 1.54$) under general instructions to intentionally fake good. Further, they hypothesized that IM and SDE increases are related more to item content and situational demands than to intentionality of distortion. Despite being equally responsive to the generic fake good instructions, IM and SDE showed substantially different effect sizes under more specific instructions. When the scale content was matched to the specific instructions, effect sizes exceeded 2.00 (compared with d s < 1.00 for mismatched pairs). The findings from these two studies question the clinical utility of distinguishing between conscious and unconscious distortion and support Rogers' (2008b) assertion that detection strategies are more effective when they match response styles. Limiting assessment to one specific response style has the added advantage of making results more interpretable (Rogers, 2008b).

Detection strategies for Axis II defensiveness. The current study investigated both general (i.e., Indiscriminate Denial of Symptoms or IDS) and focused (i.e., Denial of Commonly Reported Symptoms or D-CRS) strategies to detect denial of patient characteristics (Rogers, 2008b). Additionally, the current study introduced a tiered-level of classification, like that found in feigning research (Rogers et al., 2010), with two cut scores of varying stringency. The less stringent score can be considered a screen, focusing on sensitivity, whereas the higher threshold acts as a clinical indicator; individuals falling between the two scores may be considered “indeterminate.” This two-pronged approach maximizes both sensitivity and PPP, as suggested by Rogers (2008b).

IDS has the dual advantages of being simple to implement and adaptable to various measures. On both the SIDP-IV and the SCID-II-PQ, IDS performed adequately in identifying potential defensive responders (i.e., sensitivity on the screen) and accurately classifying defensive responders (i.e., PPP on the indicator). Because it can easily be incorporated into self-report measures, IDS has the ability to add detection strategies to the initial screens used in Axis II assessments.

Detection using D-CRS requires more information about symptom prevalence and population characteristics, but it is able to capture individuals who utilize a more subtle approach to denial. With its highly focused strategy, D-CRS produced both high sensitivity and overall accuracy (see Table 19). Moreover, the D-CRS indicator remained more accurate than other strategies even at the highest rates of prevalence, where other strategies’ accuracy diminished (see Table 22 for a comparison of detection strategies). Under more realistic estimates of defensive prevalence (i.e., 25%; Baer & Miller, 2002), D-CRS was again the most successful in achieving overall accuracy.

Table 22

Utility of Validity Scales on Four Axis II Measures in Clinical Samples

Cut Scores	Sens.	Spec.	25% Base Rate			50% Base Rate			75% Base Rate		
			PPP	NPP	OCC	PPP	NPP	OCC	PPP	NPP	OCC
MCMI-II^a/MCMI-III^b											
Y Base rate > 74 ^a	.19	.98	.80	.79	.79	.92	.55	.59	.97	.29	.39
Y Base rate ≥ 74 ^b	.60	.78	--	--	--	.73	.66	.69	--	--	--
Y Base rate ≥ 85 ^b	.35	.93	--	--	--	.82	.59	.64	--	--	--
PDS^c											
IM ≥ 9	.65	.83	.56	.88	.78	.79	.70	.74	.92	.44	.70
IM ≥ 13	.35	.98	.86	.82	.82	.95	.60	.66	.98	.33	.51
SCID-II-PQ^c											
IDS Screen < 25%	.64	.85	.58	.87	.79	.81	.70	.74	.93	.44	.69
IDS Indicator < 11%	.19	.98	.77	.79	.78	.91	.55	.59	.97	.29	.39
SIDP-IV^c											
IDS Screen < 15%	.81	.90	.74	.93	.88	.89	.83	.86	.96	.61	.83
IDS Indicator < 5%	.25	1.00	1.00	.80	.81	1.00	.57	.63	1.00	.31	.44
D-CRS Screen ≤ 4	.85	.85	.65	.94	.85	.85	.85	.85	.94	.65	.85
D-CRS Indicator ≤ 1	.42	.98	.88	.84	.84	.96	.63	.70	.96	.36	.56

Note. Sens. = sensitivity; Spec. = specificity; PPP = positive predictive power; NPP = negative predictive power; OCC = overall correct classification; IM = impression management; IDS = indiscriminate denial of symptoms; D-CRS = commonly reported symptoms.

^aFals-Stewart (1995), MCMI-II.

^bDaubert & Metzler (2000), MCMI-III. Estimates for 25% and 75% base rate could not be completed because actual classification rates were not provided. Not included in Tables 2 or 21 because diagnostic scale scores were not provided.

^cCurrent study.

Limitations

In light of Paulhus (1998), an unexpected finding was the low internal consistency seen on the PDS SDE scale. While reducing the power of the current statistical analyses, it raises more general questions regarding the homogeneity of this scale and its interpretability with severely impaired inpatient populations.

Dually diagnosed inpatients were chosen because of their high prevalence (Bakken et al., 2007; Cacciola et al., 2001; Damen et al., 2004; Jahng et al., 2004; Verheul et al., 2000) of Axis II disorders. Although necessary to exclude floor effects, exclusion of individuals with subthreshold personality pathology affects the generalizability of some current findings. In

particular, the newly created detection strategies apply only to patients with substantial Axis II pathology. As such, the specific cut scores for indiscriminate and D-CRS strategies, and likely even the content of the D-CRS scale, would need to be modified in a more diverse population. Prior to advocating the clinical use of these detection strategies, cross-validation in diverse samples with varying levels of Axis II pathology is necessary.

Replication to more diverse samples is also important for understanding the differential effect of defensive responding on specific clusters and personality disorders. The current sample's level of Axis II symptomatology was heavily weighted toward Cluster B. Although this accurately reflected other clinical samples (Damen et al., 2004; Jahng et al., 2004; Verheul et al., 2000), the effects of defensive responding on Clusters A and C deserve further examination. This is especially important for Cluster C, as previous MCMI-II studies (Fals-Stewart, 1995; Retzlaff et al., 1991) evidenced increases on Dependent and Obsessive-Compulsive Personality Disorder scales.

The current study was the first to evaluate the effects of defensive responding on an Axis II diagnostic interview. The SIDP-IV (Pfohl et al., 1997) is atypical of most diagnostic interviews in that its questions are grouped thematically, rather than by diagnostic category, and items were designed to minimize the effects of socially desirable responding. With several other interviews available, this line of research deserves expansion to determine whether other interviews are equally susceptible, especially those with more typical construction and direct questions. Regardless of test construction, the detection strategies used in the current study could easily be adapted for other interviews.

The primary constraint of the current study, and of all simulation studies, is their applicability to real-world situations. The style of responding elicited in this highly structured

setting may not directly reflect what actually occurs in diagnostic assessment. Several decisions were made in designing the current study to minimize this difference and increase ecological validity. As Rogers (2008d) and Gillard (2010) have noted, using a clinical sample, using a realistic and relevant scenario, instructing participants to be believable, and assessing understanding of and adherence to instructions all serve to enhance ecological validity. Additionally, while defensive responding was the focus of the current study due to its clinical relevance, the simulation scenario was purposefully designed to avoid specific instructions, allowing the inpatients to formulate their own approach in responding based upon the provided situation. This design is most ecologically valid, providing respondents with greater flexibility and the option of employing more than one response style at various points during an assessment. Inpatients in the current study chose to both decrease their Axis II psychopathology (i.e., defensiveness) and increase their social desirability (i.e., impression management).

The upcoming publication of *DSM-5* (APA, 2011) and its shift toward dimensional consideration of personality disorder diagnosis further limits the long-term applicability of the current study, most notably the use of diagnostic measures based on *DSM-IV* criteria. New diagnostic measures will be published in coming years based on entirely different classification system. Regardless of the diagnostic taxonomy, one lesson from the current study applies to both categorical and dimensional models of assessment: due to its ability to be successfully concealed, validity scales are necessary when measuring Axis II pathology.

Implications and Future Directions

The assessment of Axis II disorders is highly important for treatment decisions and recommendations, even when other disorders are the primary reason for treatment (Bakken et al., 2007; Cacciola et al., 2001; Hayward et al., 2006; Singer, 2005). The current thesis sought to

build upon existing simulation research in testing the susceptibility of Axis II measures to defensive responding. The results of the current study, along with findings from Bagby and Pajouhandeh (1997), Fals-Stewart (1995), and Retzlaff et al. (1991), demonstrate convincingly that personality disorder symptomatology is able to be effectively concealed on self-report measures; this susceptibility to defensive responding has now been replicated with three measures in populations with different levels of pathology. In the face of countervailing beliefs (see Widiger & Axelrod, 1995), the current study extended this finding of susceptibility to interview-based measures. The upcoming publication of the *DSM-5* and the impending development of new diagnostic measures for personality disorders make this a great time to explore in more detail the susceptibility of Axis II measures and refine methods to diminish that susceptibility or identify anomalous response styles.

By closely comparing previous models of favorable self-presentation, the current thesis also sought to clarify these assorted response styles and create detection strategies for one style in particular. Matching detection strategies to response styles (Rogers, 2008b) aids in interpretation of findings, enhances their utility, and can clarify the construct validity of existing detection strategies (Pauls & Crost, 2004). One way future simulation research can improve applicability is to use scenarios and instructions that are highly relevant to the population in question (Gillard, 2010). Relevant scenarios, tailored to each population, have the added benefit of making responses more believable (Gillard, 2010), and thus more ecologically valid.

The detection strategies created in the current study were moderately successful in identifying defensive responders. However, this prevalence-based approach would not benefit individual clinicians unless they had access to such information for their clientele. Further research on Axis II diagnostic measures ought to report average rates of endorsement for

symptoms, disorders, and total scores, making the indiscriminate denial and commonly reported symptoms strategies readily available to professionals. This data can also be provided in new or revised test manuals, based upon the normative samples.

Defensive responding is likely to reflect not only the situation but each individual's views on what is and is not desirable or possibly stigmatizing. In addition to using post hoc score corrections or indicators of invalidity, such as those developed and tested in the current investigation, researchers might consider a priori methods for dealing with social desirability. This has been dealt with indirectly by using carefully worded questions to avoid the implication of stigma or undesirability (Pfohl et al., 1997), but research has not assessed the desirability of specific symptoms like it has with other personality characteristics (e.g., Edwards, 1953). Using a clinical population to assess the desirability or stigma of each symptom (or test item) would help formulate new avenues for the detection of defensive responding on personality disorder measures. Assessing desirability of traits on a continuum (Kuncel & Tellegen, 2009) is particularly relevant given the new dimensional model of personality disorders (APA, 2011). This step would assist in creating specific, effective detection strategies for the new *DSM-5* diagnostic measures.

APPENDIX A
APPROVAL LETTER



To: Colin A. Ross, M.D.
Richard Rogers, Ph.D.
From: Craig Nuckles
CEO, Timberlawn
Re: Chelsea Fiduccia Research Project
Impression Management and the Assessment of Personality Disorders
Date: August 26, 2008

This is to confirm that the above research project was approved today by the Medical Staff Committee of Timberlawn Hospital, which acts as the Ethics Committee for the hospital. The Committee has a summary of the study and a copy of the consent form to be used in the study.

The Medical Staff understands that this project involves the administration of a set of research measures to inpatients at the hospital. These include: the Structured Interview for DSM-IV Personality; the Structured Clinical Interview for DSM-IV Personality Disorders – Personality Questionnaire; and the Paulhus Deception Scales. In addition, the researcher will review the medical chart to confirm demographic data and the clinical diagnoses.

The Medical Staff understands that participation in the study is voluntary, that non-participation will have no effect at treatment, and that patients are free to withdraw from the study at any time. No scans, blood tests or other physical procedures are involved.

The Medical Staff understands that the set of measures will be administered twice, once when patients are asked to respond honestly, and a second time when they are asked to respond in a socially desirable fashion that minimizes their personality disorder symptoms. The purpose of the project is to understand better how inpatients with Axis II personality disorders fail to disclose, to themselves and to clinicians, their Axis II symptoms, and which measures provide the most useful information about this process.

A handwritten signature in black ink, appearing to read "Craig Nuckles" followed by a date "8/26/08".

4600 Samuell Blvd. • Dallas, Texas 75228
214-381-7181 • Fax 214-388-6453

APPENDIX B
INFORMED CONSENT

University of North Texas Institutional Review Board

Informed Consent Form

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose and benefits of the study and how it will be conducted.

Title of Study: Impression Management and the Assessment of Personality Disorders

Principal Investigators: Chelsea Fiduccia, a graduate student in the University of North Texas (UNT) Department of Psychology.

Purpose of the Study: You are being asked to participate in a research study that focuses on how certain styles of responding affect the reporting of personality characteristics.

Study Procedures: Through a variety of interviews and questionnaires, you will be asked to answer questions about your personality. In Phase I, participants will be asked to respond honestly to all questions. Later, in Phase II, some participants will be asked to respond to the questions as if they were in an alternative scenario. Completion of both phases may take up to 4 hours. If you choose to participate in the “Assessment of Response Styles in Comorbid Clinical Populations” study, the information you provide may be used in conjunction with the information you provide today. This combined data may be used for additional research. Finally, your hospital records will be reviewed to gather diagnoses and demographic information.

Foreseeable Risks: The foreseeable risks are negligible. It is possible that you may find some questions minimally distressful. Please let the researcher know if this happens.

Benefits to the Subjects or Others: This study will likely help us understand how people in clinical settings respond in certain situations, thus it may help psychologists achieve effective treatment practices. Furthermore, you may gain insight about yourself from this research.

Compensation for Participants: Upon completion of both phases, you will receive \$10 as compensation for your participation. No partial compensation will be provided.

Procedures for Maintaining Confidentiality of Research Records: The confidentiality of your individual information will be maintained in any publications or presentations regarding this study. Your information will be recorded without names or other identifiers. You agree that researchers can contact the clinical staff if you pose a significant risk of suicide, self-harm, or harming others. Your name will only appear on this consent form and a record of payment, which will both be filed separately from the rest of the study documents and will in no way be linked to your information.

Questions about the Study: If you have any questions about the study, you may contact Chelsea Fiduccia, a graduate student in the University of North Texas (UNT) Department of Psychology at telephone number (940) 565-2671.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights:

Your signature below indicates that you have read or have had read to you all of the above and that you confirm all of the following:

- A graduate student researcher, Chelsea Fiduccia or Chelsea Wooley, has explained the study to you and answered all of your questions. You have been told the possible benefits and the potential risks and/or discomforts of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You have been told you will receive a copy of this form.
- I understand these records include drug/alcohol/mental health/communicable disease-related information. I understand that information released could contain reference to results of HIV antibody testing. A photocopy of this authorization should be considered as valid as the original. This consent is subject to revocation by the undersigned at any time, except to the extent that action has been taken in reliance hereon and in any event **shall expire within 90 (ninety) days** from the date of signature. The information being authorized to release is being disclosed to you from records protected by Federal confidentiality rules (42 CFR Part 2). A general authorization for the release of medical or other information is not sufficient for this purpose. The information to be released is **PRIVILEGED** and **CONFIDENTIAL** and is intended **ONLY** for research purposes.

Printed Name of Participant

Signature of Participant

Date

For the Principal Investigator or Designee:

I certify that I have reviewed the contents of this form with the subject signing above. I have explained the possible benefits and the potential risks and/or discomforts of the study. It is my opinion that the participant understood the explanation.

Signature of Principal Investigator or Designee

Date

APPENDIX C
FILE REVIEW

Identification Number: _____

Unit: Trauma Dual Diagnosis

Age: _____

Gender: Male Female

Ethnicity: European American African American

 Latin American Asian American

 Other: _____

Marital Status: Single Married Divorced Widowed

SES: Lower Middle Upper

Education Completed: Special Ed. Elementary ____ Middle ____ High ____

 GED Some college _____ Bachelors

 Masters Doctorate Other _____

Occupation: Full-time Part-time Disabled Unemployed

Axis I: _____

Axis II: _____

Axis III: _____

Axis IV: _____

Axis V: GAF Admission _____ GAF Past Year _____

APPENDIX D
SCENARIO INSTRUCTIONS

Honest:

Today I am going to ask you some questions about your personality style. Remember that your answers are confidential and will not be shared with hospital staff. The questions we will be going over today are used to help people get specialized care and treatment based on their personality. Your participation will help researchers to test how well the questionnaires work. It is very important that you answer all questions completely honestly for the most accurate picture of yourself. Can you be completely honest and accurate?

Defensive:

You've already answered questions honestly about your personality style. You will be asked everything again, but this time you are to answer based upon a pretend situation. Remember that your answers are confidential and will not be shared with hospital staff. It is very important that you answer all questions as if you were in the following situation:

You are about to be admitted to the hospital, and you've already figured out that there are two very different units. Patients on the East Unit seem to have fewer problems and are given more freedom. Patients on the West Unit seem to have severe personality problems and have trouble getting along with others.

While you realize you have some personality problems and might be a better match for the West Unit, you really want to get into the East Unit. Some of the perks of the East Unit are private bedrooms and personal TVs.

You will be meeting with another researcher in the next few minutes to complete some questionnaires. This evaluation will determine whether you go to the desirable East Unit or the more restrictive West Unit. Although you don't know exactly what it takes to

get on the East Unit, try to convince the researcher that you can get along with almost everyone and have no serious personality problems.

Warning: If you try to look perfect, the researcher won't believe you and you will lose out. Can you convince the researcher you belong on the East Unit? Can you do it without getting caught?

APPENDIX E
MANIPULATION CHECKS

Honest Condition

Identification Number: _____

1. What was the goal listed in the instructions?
 - a. Help researchers see if measures work (personalized care & treatment):
___ Correct, ___ Incorrect
 - b. Other: _____
2. How were you supposed to reach that goal? ___ Correct, ___ Incorrect, ___ Questionable
(record verbatim; answer is *be honest* or something similar):
3. Did you follow the instructions? ___ Yes, ___ No
4. (*If yes*) Which of the following (*hand participant the card*) best describes your effort today?
 - a. Didn't try hard, it's just a study ___
 - b. Tried a little bit ___
 - c. Gave a medium effort ___
 - d. A good effort, I tried hard ___
 - e. Excellent effort, I really tried to do my best ___
5. Were you successful at responding honestly?
___ Uncertain, ___ No, ___ Somewhat, ___ Yes

Defensive Condition

Identification Number: _____

1. What was the goal listed in the instructions?
 - a. Get onto East Unit (private room, TV, freedom): __ Correct, __ Incorrect
 - b. Other: _____

2. How were you supposed to reach that goal? __ Correct, __ Incorrect, __ Questionable
(record verbatim; answer is *hide personality problems* or something similar):

3. Did you follow the instructions? __ Yes, __ No

4. (*If yes*) Which of the following (*hand participant the card*) best describes your effort today?
 - a. Didn't try hard, it's just a study __
 - b. Tried a little bit __
 - c. Gave a medium effort __
 - d. A good effort, I tried hard __
 - e. Excellent effort, I really tried to do my best __

5. Do you think you were you successful at convincing me you belong on the East Unit?
__ Uncertain, __ No, __ Somewhat, __ Yes

6. What did you do to try to get onto the East Unit?

REFERENCES

- Amador, X. F., Strauss, D. H., Yale, S. A., Flaum, M. M., Endicott, J., & Gorman, J. M. (1993). Assessment of insight in psychosis. *American Journal of Psychiatry, 150*, 873-879.
Retrieved from <http://ajp.psychiatryonline.org/cgi/reprint/150/6/873>
- American Law Institute (1962). *Model penal code*. Philadelphia: Author.
- American Psychiatric Association. (1952). *Diagnostic and statistical manual: Mental disorders*. Washington, DC: Author.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- American Psychiatric Association. (2011, June 21). *Changes to the reformulation of personality disorders for DSM-5*. Retrieved from <http://www.dsm5.org/proposedrevision/Pages/PersonalityDisorders.aspx>
- Aviram, R. B., Brodsky, B. S., & Stanley, B. (2006). Borderline personality disorder, stigma, and treatment implications. *Harvard Review of Psychiatry, 14*(5), 249-256.
doi:10.1080/10673220600975121
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14*(1), 16-26. doi: 10.1037//1040-3590.14.1.16
- Bagby, R., Buis, T., & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: Replication and extension. *Psychological Assessment, 7*(1), 84-92. doi:10.1037/1040-3590.7.1.84

- Bagby, R., Gillis, J., Toner, B. B., & Goldberg, J. (1991). Detecting fake-good and fake-bad responding on the Millon Clinical Multiaxial Inventory-II. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(3), 496-498. doi:10.1037/1040-3590.3.3.496
- Bagby, R., & Pajouhandeh, P. (1997). The detection of faking good on the Personality Diagnostic Questionnaire-4. *Assessment*, 4(3), 305-309. Retrieved from <http://asm.sagepub.com>
- Bagby, R., Rogers, R., Nicholson, R. A., Buis, T., Seeman, M. V., & Rector, N. A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of defensive responding in clinical and nonclinical samples. *Psychological Assessment*, 9(4), 406-413. doi:10.1037/1040-3590.9.4.406
- Bakken, K., Landheim, A., & Vaglum, P. (2007). Axis I and II disorders as long-term predictors of mental distress: A six-year prospective follow-up of substance-dependent patients. *BMC Psychiatry*, 7(29). doi:10.1186/1471-244X-7-29
- Ball, S. A., Rounsaville, B. J., Tennen, H., & Kranzler, H. R. (2001). Reliability of personality disorder symptoms and personality traits in substance-dependent inpatients. *Journal of Abnormal Psychology*, 110(2), 341-352. doi:10.1037/0021-843X.110.2.341
- Beck, J. C. (2010). Dangerous severe personality disorder: The controversy continues. *Behavioral Sciences & the Law*, 28(2), 277-288. doi: 10.1002/bsl.931
- Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlasgan, T. H., & ... Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of Psychiatry*, 158(2), 295-302. doi:10.1176/appi.ajp.158.2.295

- Borge, F.-M., Hoffart, A., Sexton, H., Martinsen, E., Gude, T., Hedley, L. M., & Abrahamsen, G. (2010). Pre-treatment predictors and in-treatment factors associated with change in avoidant and dependent personality disorder traits among patients with social phobia. *Clinical Psychology & Psychotherapy, 17*(2), 87-99. doi: 10.1002/cpp.640
- Buller, D. B., Strzyzewski, K. D., & Hunsaker, F. G. (1991). Interpersonal deception: II. The inferiority of conversational participants as deception detectors. *Communication Monographs, 58*(1), 25-40. doi:10.1080/03637759109376212
- Butcher, J. N. (2010). Personality assessment from the nineteenth to the early twenty-first century: Past achievements and contemporary challenges. *Annual Review of Clinical Psychology, 61*-20. doi:10.1146/annurev.clinpsy.121208.131420
- Cacciola, J. S., Alterman, A. I., McKay, J. R., & Rutherford, M. J. (2001). Psychiatric comorbidity in patients with substance use disorders: Do not forget Axis II disorders. *Psychiatric Annals, 31*(5), 321-331. Retrieved from <http://www.psychiatricannalsonline.com/bissues.asp> Cacciola, Alterman, McKay, & Rutherford, 2001
- Charland, L. C. (2006). Moral nature of the DSM-IV cluster B personality disorders. *Journal of Personality Disorders, 20*(2), 116-125. doi:10.1521/pedi.2006.20.2.116
- Clark, L. A. (2006). The role of moral judgment in personality disorder diagnoses. [Introduction to special issue]. *Journal of Personality Disorders, 20*(2), 184-185. doi:10.1521/pedi.2006.20.2.184
- Clark, L. A., Simms, L. J., Wu, K. D., & Casillas, A. (in press). *Schedule for Nonadaptive and Adaptive Personality—Second Edition (SNAP-2)*. Minneapolis, MN: University of Minnesota Press.

- Coolidge, F. L., & Merwin, M. M. (1992). Reliability and validity of the Coolidge Axis II Inventory: A new inventory for the assessment of personality disorders. *Journal of Personality Assessment*, *59*(2), 223-238. doi:10.1207/s15327752jpa5902_1
- Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7) (614-625). doi:10.1037/0003-066X.59.7.614
- Corrigan, P. W., Markowitz, F. E., & Watson, A. C. (2004). Structural levels of mental illness stigma and discrimination. *Schizophrenia Bulletin*, *30*(3), 481-491. Retrieved from <http://schizophreniabulletin.oxfordjournals.org/content/30/3/481.full.pdf+html>
- Damen, K. M., De Jong, C. J., & Van der Kroft, P. A. (2004). Interrater reliability of the Structured Interview for DSM-IV Personality in an opioid-dependent patient sample. *European Addiction Research*, *10*(3), 99-104. doi:10.1159/000077697
- Daubert, S. D., & Metzler, A. E. (2000). The detection of fake-bad and fake-good responding on the Millon Clinical Multiaxial Inventory III. *Psychological Assessment*, *12*(4), 418-424. doi:10.1037/1040-3590.12.4.418
- Djikic, M., Peterson, J. B., & Zelazo, P. (2005). Attentional biases and memory distortions in self-enhancers. *Personality and Individual Differences*, *38*(3), 559-568. doi:10.1016/j.paid.2004.05.010
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, *37*(2), 90-93. doi:10.1037/h0058073
- Exner, J. E. (2003). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations and principles of interpretation* (4th ed.). New York: Wiley.

- Fals-Stewart, W. (1995). The effect of defensive responding by substance-abusing patients on the Million Clinical Multiaxial Inventory. *Journal of Personality Assessment*, *64*, 540-551. doi:10.1207/s15327752jpa6403_11
- Farmer, R. F. (2000). Issues in the assessment and conceptualization of personality disorders. *Clinical Psychology Review*, *20*, 823-851. doi:10.1016/S0272-7358(99)00014-8
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, D.C.: American Psychiatric Press, Inc.
- Ganellen, R. J. (2007). Assessing normal and abnormal personality functioning: Strengths and weaknesses of self-report, observer, and performance-based methods. *Journal of Personality Assessment*, *89*(1), 30-40. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/00223890701356987>
- Germans, S., Van Heck, G. L., Moran, P., & Hodiament, P. G. (2008). The self-report standardized assessment of personality-abbreviated scale: Preliminary results of a brief screening test for personality disorders. *Personality and Mental Health*, *2*(2), 70-76. doi:10.1002/pmh.34
- Gillard, N. D. (2010). *Methodological issues in malingering research: The use of simulation designs* (master's thesis, University of North Texas). Retrieved from <http://digital.library.unt.edu/ark:/67531/metadc33153/>
- Grant B. F., Dawson D. A., & Hasin D. S. (2001). *The Alcohol Use Disorders and Associated Disabilities Interview Schedule—Version for DSM-IV (AUDADIS-IV)*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

- Grant, B. F., Hasin, D. S., Stinson, F. S., Dawson, D. A., Chou, S. P., Ruan, W. J., & Pickering, R. P. (2004). Prevalence, correlates, and disability of personality disorders in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry, 65*(7), 948-958. Retrieved from <http://article.psychiatrist.com/?ContentType=START&ID=10000967>
- Greene 2000 (MMPI book) Greene, R. L. (2000). *The MMPI-2: An interpretive manual*. Boston, MA: Allyn and Bacon.
- Hathaway, S. R., & McKinley, J. C. (1940). *The MMPI manual*. New York: Psychological Corporation.
- Hayward, M., Slade, M., & Moran, P. A. (2006). Personality disorders and unmet needs among psychiatric inpatients. *Psychiatric Services, 57*, 538-543. doi:10.1176/appi.ps.57.4.538
- Helmes, E. (2009). Conventional and actuarial methods to detect response distortion on the basic personality inventory. *Canadian Journal of Behavioural Science, 41*(3), 121-132. doi:10.1037/a0015589
- Holden, R. R., & Jackson, D. N. (1979). Item subtlety and face validity in personality assessment. *Journal of Consulting and Clinical Psychology, 47*(3), 459-468. doi:10.1037/0022-006X.47.3.459
- Holden, R. R., Starzyk, K. B., McLeod, L. D. & Edwards, M. J. (2000). Comparisons among the Holden Psychological Screening Inventory (HPSI), the Brief Symptom Inventory (BSI), and the Balanced Inventory of Desirable Responding (BIDR). *Assessment, 7*(2), 163-175. doi:10.1177/107319110000700208
- Hyler, S. E. (1994). *PDQ-4 Personality Questionnaire*. Unpublished scale. New York: New York State Psychiatric Institute.

- Jackson, D. N. (1996). *Basic Personality Inventory manual* (2nd ed.). London, Canada: Sigma Assessment Systems.
- Jahng, S., Trull, T. J., Wood, P. K., Tragesser, S. L., Tomko, R., Grant, J. D., & ... Sher, K. J. (2011). Distinguishing general and specific personality disorder features and implications for substance dependence comorbidity. *Journal of Abnormal Psychology*. Advance online publication. doi:10.1037/a0023539
- Jane, J. S., Pagan, J. L., Turkheimer, E., Fiedler, E. R., & Oltmanns, T. F. The interrater reliability of the Structured Interview for DSM-IV Personality. *Comprehensive Psychiatry*, 47, 368-375. doi:10.1016/j.comppsy.2006.01.009
- Kendell, R. E. (2002). The distinction between personality disorder and mental illness. *British Journal of Psychiatry*, 180(2), 110-115. doi:10.1192/bjp.180.2.110
- Klein, D. N. (2003). Patients' versus informant reports of personality disorders in predicting 7½-year outcome in outpatients with depressive disorders. *Psychological Assessment*, 15, 216-222. doi:10.1037/1040-3590.15.2.216
- Klein, M. H., & Benjamin, L. S. (1996). *The Wisconsin Personality Disorders Inventory-IV*. Madison, WI: University of Wisconsin.
- Klonsky, E., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice*, 9(3), 300-311. doi:10.1093/clipsy/9.3.300
- Kuncel, N. R., & Telegen, A. (2009). A conceptual and empirical reexamination of the measurement of the social desirability of items: Implications for detecting desirable response style and scale development. *Personnel Psychology*, 62(2), 201-228. doi:10.1111/j.1744-6570.2009.01136.x

- Langbehn, D. R., Pfohl, B. M., Reynolds, S., Clark, L., Battaglia, M., Bellodi, L., & ... Links, P. (1999). The Iowa Personality Disorder Screen: Development and preliminary validation of a brief screening interview. *Journal of Personality Disorders, 13*(1), 75-89. Retrieved from <http://www.psych.umn.edu/faculty/grove/093iowapersonalitydisorderscreen.pdf>
- Lanyon, R. I. (2001). Dimensions of self-serving misrepresentation in forensic assessment. *Journal of Personality Assessment, 76*(1), 169-179. doi:10.1207/S15327752JPA7601_10
- Lanyon, R. I. (2004). Favorable self-presentation on psychological inventories: An analysis. *American Journal of Forensic Psychology, 22*(1), 53-65. Retrieved from <http://www.forensicpsychology.org/journal.htm>
- Lanyon, R. I. (2007). Utility of the Psychological Screening Inventory: A review. *Journal of Clinical Psychology, 63*, 283-307. doi:10.1002/jclp.20344
- Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity of scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement, 67*, 859-876. doi:10.1177/0013164406299104
- Leichsenring, F., & Leibing, E. (2003). The effectiveness of psychodynamic therapy and cognitive behavior therapy in the treatment of personality disorders: A meta-analysis. *American Journal of Psychiatry, 160*(7), 1223-1232. Retrieved from <http://ajp.psychiatryonline.org/cgi/reprint/160/7/1223>
- Lenzenwenger, M. F., Lane, M. C., Loranger, A. W., & Kessler, R. C. (2007). DSM-IV personality disorders in the national comorbidity survey replication. *Biological Psychiatry, 62*, 553-564. doi:10.1016/j.biopsych.2006.09.019
- Lewis, G., & Appleby, L. (1988). Personality disorder: The patients psychiatrists dislike. *British Journal of Psychiatry, 153*, 44-49. doi:10.1192/bjp.153.1.44

- Li, A. & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement, 67*(3), 525-544. doi:10.1177/0013164406292087
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*(2), 27-66. doi:10.1111/1529-1006.002
- Lincoln, T. M., Lüllmann, E., & Rief, W. (2007). Correlates and long-term consequences of poor insight in patients with schizophrenia: A systematic review. *Schizophrenia Bulletin, 33*(6), 1324-1342. doi:10.1093/schbul/sbm002
- Link, B. G., & Phelan, J. C. (2001). Conceptualizing stigma. *Annual Review of Sociology, 27*, 363-385. doi:10.1146/annurev.soc.27.1.363
- Link, B. G., & Phelan, J. C. (2006). Stigma and its public health implications. *Lancet, 367*(9509), 528-529. doi:10.1016/S0140-6736(06)68184-1
- Loranger, A. W. (1988). *Personality Disorder Examination (PDE) manual*. Yonkers, New York: DV Communications.
- Loranger, A. W. (1999). *The International Personality Disorder Examination (IPDE) DSM-IV and ICD-10 Modules*. Odessa, FL: Psychological Assessment Resources.
- Loranger, A. W. (2002). *OMNI personality inventory and OMNI-IV personality disorder inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Luborsky, L., Auerbach, A. H., Chandler, M., Cohen, J., & Bachrach, H. M. (1971). Factors influencing the outcome of psychotherapy: A review of quantitative research. *Psychological Bulletin, 75*(3), 145-185. doi:10.1037/h0030480

- Mann, A. H., Jenkins, R. R., Cutting, J. C., & Cowen, P. J. (1981). The development and use of a standardized assessment of abnormal personality. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences*, *11*(4), 839-847.
doi:10.1017/S0033291700041337
- Matusiewicz, A. K., Hopwood, C. J., Banducci, A. N., & Lejuez, C. W. (2010). The effectiveness of cognitive behavioral therapy for personality disorders. *Psychiatric Clinics of North America*, *33*(3), 657-685. doi:10.1016/j.psc.2010.04.007
- McCann, J. T., Flens, J. R., Campagna, V., Collman, P., Lazzaro, T., & Connor, E. (2001). The MCMI-III in child custody evaluations: A normative study. *Journal of Forensic Psychology Practice*, *1*(2), 27-44. doi:10.1300/J158v01n02_02
- Meehl, P. E. (1945). The dynamics of 'structured' personality tests. *Journal of Clinical Psychology*, *1*, 296-303.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*(5), 525-564.
doi:10.1037/h0053634
- Meyer, G. J., & Kurtz, J. E. (2006). Advancing personality assessment terminology: Time to retire 'objective' and 'projective' as personality test descriptors. *Journal of Personality Assessment*, *87*(3), 223-225. doi:10.1207/s15327752jpa8703_01
- Miller, J. D., Bagby, R. M., Pilkonis, P. A., Reynolds, S. K., & Lynam, D. R. (2005). A simplified technique for scoring DSM-IV personality disorders with the five-factor model. *Assessment*, *12*, 404-415. doi:10.1177/1073191105280987
- Millon, T., Davis, R., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory—III manual* (2nd ed.). Minneapolis, MN: Pearson.

- Mohino, S., Pujol, A., & Idiaquez, I. (2011). Personality disorders and criminal responsibility in the Spanish Supreme Court. *Journal of Forensic Sciences*, 56(1), 150-154.
doi:10.1111/j.1556-4029.2010.01558.x
- Moran, P., Leese, M., Lee, T., Walters, P., Thornicroft, G., & Mann, A. (2003). Standardised assessment of personality - abbreviated scale (SAPAS): Preliminary validation of a brief screen for personality disorder. *British Journal of Psychiatry*, 183(3), 228-232.
doi:10.1192/bjp.183.3.228
- Oldham, J. M. (2005). Personality disorders. *Journal of Lifelong Learning in Psychiatry*, 3(3), 372-382. Retrieved from <http://focus.psychiatryonline.org/cgi/reprint/3/3/372>
- Oltmanns, T. F., Gleason, M. J., Klonsky, E., & Turkheimer, E. (2005). Meta-perception for pathological personality traits: Do we know when others think that we are difficult? *Consciousness and Cognition: An International Journal*, 14(4), 739-751.
doi:10.1016/j.concog.2005.07.001
- Overton, S. L. & Medrina, S. L. (2008). The stigma of mental illness. *Journal of Counseling and Development*, 86, 143-151. Retrieved from <http://aca.metapress.com/openurl.asp?genre=article&eissn=1556-6676&volume=86&issue=2&spage=143>
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598-609. doi:10.1037/0022-3514.46.3.598
- Paulhus, D. L. (1998). *Paulhus Deception Scales (PDS): The Balanced Inventory of Desirable Responding-7, user's manual*. North Tonawanda, NY: Multi-Health Systems.
- Paulhus, D. L. (1999, April). *BIDR Self-Deceptive Denial (SDD)*. Retrieved from <http://neuron4.psych.ubc.ca/~dpaulhus/research/SDR/downloads/MEASURES/SDD.htm>.

- Pauls, C. A. & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences*, 37(6), 1137-1151.
doi:10.1016/j.paid.2003.11.018
- Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured Interview for DSM-IV Personality*. Washington, DC: American Psychiatric Press.
- Pilkonis, P. A., & Frank, E. (1988). Personality pathology in recurrent depression: Nature, prevalence, and relationship to treatment response. *American Journal of Psychiatry*, 145, 435-441.
- Retzlaff, P., Sheehan, E., & Fiel, A. (1991). MCMI-II report style and bias: Profile and validity scales analyses. *Journal of Personality Assessment*, 56, 466-477.
doi:10.1207/s15327752jpa5603_8
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, 21(2), 95-103. doi:10.1037/h0045357
- 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 Press.
- Rogers, R. (2008b). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14-35). New York: Guilford Press.
- Rogers, R. (2008c). Structured interviews and dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 301-321). New York: Guilford Press.
- Rogers, R. (2008d). Researching response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 411-434). New York: Guilford Press.

- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms* (2nd ed., Professional manual). Odessa, FL: Psychological Assessment Resources.
- Rush, B., & Koegl, C. J. (2008). Prevalence and profile of people with co-occurring mental and substance use disorders within a comprehensive mental health system. *Canadian Journal of Psychiatry, 53*(12), 810-821. Retrieved from <http://publications.cpa-apc.org/media.php?mid=719>
- Salekin, R.T. (2000). Test review: The Paulhus Deception Scales [Review of the Paulhus Deception Scales]. *AP-LS News, Fall 2000*, 8-11. Retrieved from <http://www.ap-ls.org/publications/newsletters/fall2000.pdf>
- Sartorius, N. (2002). Iatrogenic stigma of mental illness [Editorial]. *BMJ, 324*, 1470-1471. doi:10.1136/bmj.324.7352.1470
- Schulze, B. (2007). Stigma and mental health professionals: A review of the evidence on an intricate relationship. *International Review of Psychiatry, 19*(2), 137-155. doi:10.1080/09540260701278929
- 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 Press.
- Sewell, K. W. (2008). Dissimulation on projective measures. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed. pp. 207-217). New York: Guilford Press.
- Singer, J. A. (2005). *Personality and psychotherapy*. New York: Guilford Press.

- Sriram, T. G. & Jabbarpour, Y. M. (2005). Are mental health professionals immune to stigmatizing beliefs? [Letter to the editor]. *Psychiatric Services*, *56*(5), 610.
doi:10.1176/appi.ps.56.5.610
- Stagner, R. R. (1934). Validity and reliability of the Bernreuter Personality Inventory. *Journal of Abnormal and Social Psychology*, *28*(4), 413-418. doi:10.1037/h0071099
- Stalker, K., Ferguson, I., & Barclay, A. (2005). 'It is a horrible term for someone': Service user and service provider perspectives on 'personality disorder.' *Disability and Society*, *20*(4), 359-373. doi:10.1080/09687590500086443
- Stöber, J., Dette, D., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment*, *78*(2), 370-389. doi:10.1207/S15327752JPA7802_10
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103.
doi:10.1207/S15327752JPA8001_18
- Triebwasser, J., & Siever, L. J. (2007). Pharmacotherapy of personality disorders. *Journal of Mental Health*, *16*(1), 5-50. doi:10.1080/09638230601182078
- Tyrer, P., Alexander, J. & Ferguson B. (1988) Personality Assessment Schedule. In P. Tyrer (Ed.) *Personality Disorders: Diagnosis, Management and Course* (pp. 140-167). London: Butterworth/Wright.
- Tyrer, P., & Mulder, R. (2006). Management of complex and severe personality disorders in community mental health services. *Current Opinion in Psychiatry*, *19*(4), 400-404.
doi:10.1097/01.yco.0000228761.02861.44

- Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5(3), 243-262. doi:10.1177/1745691610369465
- Van Beek, N., & Verheul, R. (2008). Motivation for treatment in patients with personality disorders. *Journal of Personality Disorders*, 22(1), 89-100. doi:10.1521/pedi.2008.22.1.89
- Verheul, R. (2005). Clinical utility of dimensional models for personality pathology. *Journal of Personality Disorders*, 19(3), 283-302. doi:10.1521/pedi.2005.19.3.283
- Verheul, R., Kranzler, H. R., Poling, J., Tennen, H., Ball, S., & Rounsaville, B. J. (2000). Co-occurrence of Axis I and Axis II disorders in substance abusers. *Acta Psychiatrica Scandinavica*, 101, 110-118. doi:10.1034/j.1600-0447.2000.90050.x
- Viswesvaran, C. & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197-210. doi:10.1177/00131649921969802
- Westen, D. (1997). Divergences between clinical and research methods for assessing personality disorders: Implications for research and the evolution of Axis II. *American Journal of Psychiatry*, 154(7), 895-903. Retrieved from <http://ajp.psychiatryonline.org/cgi/content/abstract/154/7/895>
- Westen, D., & Shedler, J. (1999). Revising and assessing axis II, part I: Developing a clinically and empirically valid assessment method. *American Journal of Psychiatry*, 156(2), 258-272. Retrieved from <http://ajp.psychiatryonline.org/archive/>

- Whyte, S., Fox, S., & Coxell, A. (2006). Reporting of personality disorder symptoms in a forensic inpatient sample: Effects of mode of assessment and response style. *Journal of Forensic Psychiatry and Psychology, 17*(3), 431-441. doi:10.1080/14789940600775436
- Widiger, T. A. (1997). Personality disorders as maladaptive variants of common personality traits: Implications for treatment. *Journal of Contemporary Psychotherapy, 27*(4), 265-282. doi:10.1023/A:1025662407740
- Widiger, T. A. (2007). Current controversies in nosology and diagnosis of personality disorders. *Psychiatric Annals, 37*(2), 93-99. Retrieved from <http://www.psychiatricannalsonline.com/view.asp?rid=20979>
- Widiger, T. A., & Axelrod, S. R. (1995). Recent developments in the clinical assessment of personality disorders. *European Journal of Psychological Assessment, 11*(3), 213-221. doi:10.1027/1015-5759.11.3.213
- Widiger, T. A., Mangine, S., Corbitt, E.M., Ellis, C. G., & Thomas, G. V. (1995). *Personality Disorder Interview-IV. Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Widiger, T. A. & Samuel, D. B. (2005). Evidence-based assessment of personality disorders. *Psychological Assessment, 17*(3), 278-287. doi:10.1037/1040-3590.17.3.278
- Wilberg, T., Hummelen, B., Pedersen, G., & Karterud, S. (2008). A study of patients with personality disorder not otherwise specified. *Comprehensive Psychiatry, 49*, 460-468. doi:10.1016/j.comppsy.2007.12.008
- Zachar, P., & Potter, N. (2010). Personality disorders: Moral or medical kinds—Or both?. *Philosophy, Psychiatry, & Psychology, 17*(2), 101-117. doi:10.1353/ppp.0.0290

Zanarini, M. C., Frankenburg, F. R., Sickel, A. E., & Young, L. (1996). *The Diagnostic Interview for DSM-IV Personality Disorders (DIPD-IV)*. Belmont, MA, McLean Hospital.

Zimmerman, M. & Mattia, J. I. (1999). Psychiatric diagnosis in clinical practice: Is comorbidity being missed? *Comprehensive Psychiatry*, 40(3), pp. 182-191. doi:10.1016/S0010-440X(99)90001-9

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, 162(10), 1911-1918. doi:10.1176/appi.ajp.162.10.1911