CRITERIA COMBINATIONS IN THE PERSONALITY DISORDERS: CHALLENGES ASSOCIATED WITH A POLYTHETIC DIAGNOSTIC SYSTEM

A Thesis

by

LUKE D. COOPER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2010

Major Subject: Psychology

CRITERIA COMBINATIONS IN THE PERSONALITY DISORDERS: CHALLENGES ASSOCIATED WITH A POLYTHETIC DIAGNOSTIC SYSTEM

A Thesis

by

LUKE D. COOPER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Stephen Balsis
Committee Members,	Leslie C. Morey
	Myeongsun Yoon
Head of Department,	Leslie C. Morey

May 2010

Major Subject: Psychology

ABSTRACT

Criteria Combinations in the Personality Disorders: Challenges Associated with a Polythetic Diagnostic System. (May 2010)

Luke D. Cooper, B.A., Rowan University

Chair of Advisory Committee: Dr. Stephen Balsis

Converging research on the diagnostic criteria for personality disorders (PDs) reveals that most criteria have different psychometric properties. This finding is inconsistent with the *DSM-IV-TR* PD diagnostic system, which weights each criterion equally. The purpose of the current study was to examine the potential effects of using equal weights for differentially-functioning criteria. Using data from over 2,100 outpatients, response patterns to the diagnostic criteria for nine PDs were analyzed and scored within an item response theory (IRT) framework. Results indicated that combinations that included the same number of endorsed criteria (the same "raw score") yielded differing estimates of PD traits, depending on which criteria were met. Moreover, trait estimates from subthreshold criteria combinations often overlapped with diagnostic combinations (i.e., at threshold or higher), indicating that there were subthreshold combinations of criteria that indicated as much or more PD traits than some combinations at the diagnostic threshold. These results suggest that counting the number of criteria an individual meets provides only a coarse estimation of their PD trait level.

Suggestions for the improved measurement of polythetically-defined mental disorders are discussed.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Steve Balsis, and the members of my thesis committee, Drs. Morey and Yoon, for their feedback and guidance in this project. I would also like to thank Dr. Mark Zimmerman for his generosity in allowing access to his data, and for his collaboration and feedback. Finally, I would like to thank my family and friends for their ongoing support. Thank you all.

NOMENCLATURE

DSM	Diagnostic and Statistical Manual of Mental Disorders
PD/PDs	Personality Disorder/Personality Disorders
IRT	Item Response Theory
SIDP-IV	Structured Interview for DSM-IV Personality
CFA/CFAs	Confirmatory Factor Analysis/Confirmatory Factor Analyses
2PL	Two-Parameter Logistic
ICC	Item Characteristic Curve
EVS	Expected Value Score
ML/MLE	Maximum Likelihood/Maximum Likelihood Estimate

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Although many algorithms are used in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision; *DSM-IV-TR*; American Psychiatric Association, 2000) to help classify and diagnose mental disorders, most can be grouped into two simple approaches. *Monothetically-defined disorders* require that every diagnostic criterion must be met, and *polythetically-defined disorders* require only that a specified number of the diagnostic criteria be met. The 10 *DSM-IV-TR* personality disorders (PDs) are fully polythetic, meaning that one needs to meet only a subset of a disorder's total number of diagnostic criteria, and that no specific criterion must be present in order to be considered for diagnosis. Although monothetic criteria sets were once used to classify four PDs in the *DSM-III*, they were found to be less reliable (Pfohl, Coryell, Zimmerman, & Stangl, 1986) and too restrictive for diagnosing these heterogeneous disorders (cf. Widiger, Frances, Spitzer, & Williams, 1988). Polythetic criteria sets, on the other hand, allow for phenotypic variation in the symptom manifestations of a disorder, thus providing more diagnostic flexibility. Perhaps as a result of these considerations, all PDs since the *DSM-III-R* have been classified polythetically.

With the adoption of polythetic criteria sets, however, additional considerations arise. For example, when only a subset of criteria is needed for a diagnosis, a cutoff must be implemented to determine when the disorder is present. For the PDs this cutoff,

---

This thesis follows the style of the *Journal of Abnormal Psychology*.

known as the diagnostic threshold, is based solely on the number of criteria met. When an individual endorses enough criteria to meet or exceed the diagnostic threshold (as well as meeting the other diagnostic requirements), he or she becomes eligible for the diagnosis (APA, 2000). Conversely, an individual cannot technically be diagnosed with one of the 10 *DSM-IV-TR* PDs without endorsing the requisite number of criteria, even if he or she meets the other diagnostic requirements.

Another important consideration in a polythetic system is that many combinations of criteria can become diagnostic (i.e., can meet or exceed the threshold). Consider a disorder with seven diagnostic criteria. If the disorder were classified monothetically, there would be only one way for an individual to be diagnosed: by meeting all seven criteria. If, however, a threshold were placed to allow any six of the seven criteria to be diagnostic (now a polythetically defined disorder), there would be eight ways an individual could be diagnosed with the disorder: seven ways to meet six of the seven criteria (e.g., all but criterion one, all but criterion two, etc.), and one way to meet all seven criteria. Further decreasing the diagnostic threshold allows more combinations to become diagnostic. If the threshold were lowered to five, for example, there would then be 29 ways to be diagnosed: 21 ways to meet any five of seven criteria plus the eight ways to meet the combinations of six and seven criteria. And so on. Disorders with more diagnostic criteria have an even greater number of possible combinations. For borderline PD, a disorder with nine diagnostic criteria and a diagnostic threshold of five, there are 256 diagnostic combinations of criteria. Table 1

displays the number of combinations for each disorder at each raw score (a raw score is equal to the number of diagnostic criteria met).[1]

The many diagnostic combinations permitted by a polythetic system are problematic (e.g., Frances, Pincus, Widiger, Davis, & First, 1990). Two individuals diagnosed with the same PD may only share one or two diagnostic features, and the thresholds for antisocial and obsessive-compulsive PD make it possible for two individuals with the same diagnosis to have no diagnostic features in common. In addition to the large amount of heterogeneity permitted by a polythetic diagnostic system, the numerous combinations of criteria present another diagnostic dilemma: Converging empirical evidence on the diagnostic criteria for the PDs indicates that all criteria are not "created equal."

Research examining the diagnostic efficiency (e.g., sensitivity, specificity, positive predictive power, and negative predictive power; see Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984) of the PD criteria has repeatedly shown that certain criteria and combinations of criteria are more useful than others in the identification and diagnosis of PD (e.g., Farmer & Chapman, 2002; Grilo et al., 2001; Pfohl et al., 1986).

---

[1] The total number of response patterns to a dichotomously scored criteria set (i.e., each criterion is either present or absent) equals 2^n, where n is the number of diagnostic criteria. The number of combinations to the eight dependent PD criteria, for example, is 2^8 or 256. Combinations of criteria at each raw score can be calculated using the combination formula, $n! / (r! (n - r)!)$, where r is a raw score, n is the total number of criteria for a given PD, and ! is the symbol for a factorial (e.g., $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$). The number of ways to meet 5 (r) of the 8 (n) histrionic diagnostic criteria, for example, is $8! / (5! \times (3)!)$, or 56. The number of combinations at each raw score within a criteria set also sum to 2^n. These equations apply to most criterion sets, not only those for the PDs.

Table 1

Criteria Combinations for the Personality Disorders at Each Raw Score

Disorder	#	Thresh	0	1	2	3	4	5	6	7	8	9	Total	Diag
Antisocial	7	3	1	7	21	**35**	**35**	**21**	**7**	**1**	-	-	128	99
Avoidant	7	4	1	7	21	35	**35**	**21**	**7**	**1**	-	-	128	64
Paranoid	7	4	1	7	21	35	**35**	**21**	**7**	**1**	-	-	128	64
Schizoid	7	4	1	7	21	35	**35**	**21**	**7**	**1**	-	-	128	64
Obs-Comp	8	4	1	8	28	56	**70**	**56**	**28**	**8**	**1**	-	256	163
Dependent	8	5	1	8	28	56	70	**56**	**28**	**8**	**1**	-	256	93
Histrionic	8	5	1	8	28	56	70	**56**	**28**	**8**	**1**	-	256	93
Borderline	9	5	1	9	36	84	126	**126**	**84**	**36**	**9**	**1**	512	256
Narcissistic	9	5	1	9	36	84	126	**126**	**84**	**36**	**9**	**1**	512	256
Schizotypal	9	5	1	9	36	84	126	**126**	**84**	**36**	**9**	**1**	512	256

Note. # = the number of diagnostic criteria; Thresh = the diagnostic threshold; Total = the total number of response patterns to the criteria; Diag = the number of possible ways to be diagnosed. Numbers in bold signify combinations that are at or above the diagnostic threshold.

Studies on the prototypicality of PD features have shown that some criteria are rated as more prototypical, or central, to each disorder concept than others (e.g., Kim & Ahn, 2002; Livesley, Reiffer, Sheldon, & West, 1987; Shedler & Westen, 2004). Research on clinicians' use of the PD diagnostic criteria has shown that clinicians tend to weight certain criteria more heavily than others (Davis, Blashfield, & McElroy, 1993; Evans, Herbert, Nelson-Gray, & Gaudiano, 2002; Kim & Ahn, 2002; Morey & Ochoa, 1989). And more recently, analyses using item response theory (IRT) have shown that the PD criteria differ in their ability to measure the latent disorder trait, with some criteria measuring higher levels of the disorder trait than others (Bienfait, 2008; Cooper & Balsis, 2009; Feske, Kirisci, Tarter, & Pilkonis, 2007).

Results from these methodologically diverse approaches all suggest that it is important to consider which PD criteria are endorsed. Yet, the criteria-counting algorithm currently used in the diagnosis of PDs essentially disregards the psychometric properties of individual criteria. And in the presence of differently-functioning criteria, counting the number of criteria met is not an ideal diagnostic algorithm, as it models only the "quantity" (but not the varying "quality") of a disorder's diagnostic criteria (Aggen, Neale, & Kendler, 2005; Clarke & McKenzie, 1994, p. 187; Sakashita, Slade, & Andrews, 2007).

Recognizing that the diagnostic criteria do not have identical psychometric properties, developers of the *DSM-IV* rank-ordered the criteria for most PDs in terms of

their diagnostic "importance" (APA, 2000; p. 686).[2] The implication again being that some criteria may be more related to, or better at detecting the presence of a diagnostic construct than others. A differential weighting scheme was considered for use in the *DSM-IV* to account for the relative contribution of each PD criterion, but developers ultimately decided to retain the traditional (unit-weighted) polythetic format for its own advantages (see Gunderson, 1998). Although there were legitimate reasons to opt for the traditional polythetic format, this approach remains at odds with the psychometric research on the PD criteria. And as a result of this inconsistency, diagnostic discrepancies may arise.

In a previous study (Cooper & Balsis, 2009), we analyzed the diagnostic criteria for schizoid PD, and calculated latent trait estimates for all possible combinations of its criteria. The results revealed that each criterion functioned differently, and that combinations of criteria with the same raw score resulted in varying estimates of the latent schizoid trait. At a raw score of four, for example, combinations that included the criteria that were most related to the construct had a greater trait estimates than any other combination of four endorsed criteria. The combinations that contained either the most or least related criteria represented the maximum and minimum trait estimates at each raw score—the remaining combinations of the same raw score falling somewhere in between. The authors termed these varying estimates of the latent trait at each raw score

[2] "Importance" was largely operationalized in terms of each criterion's diagnostic efficiency characteristics, with specific attention paid to phi coefficients and specificity values (Gunderson, 1998).

bands; each band representing the range of the latent trait to which that raw score corresponded.

Of further note, the bands from one raw score overlapped with the bands from abutting raw scores, indicating that an individual can endorse fewer diagnostic criteria, yet have as much of the disorder trait as an individual who endorses more criteria. There were even subthreshold combinations of the schizoid criteria that had higher trait estimates than the estimates from several at-threshold combinations. Indeed, an individual could meet a subthreshold number of diagnostic criteria (which would fail to meet the *DSM-IV-TR* requirement for a diagnosis), yet have as much (or more!) of the schizoid trait than some individuals who met the threshold for diagnosis (Cooper & Balsis, 2009).

These results suggested that using criteria counts in the assessment of polythetically-defined mental disorders may not be an accurate measurement procedure, and that diagnostic thresholds based on the number of criteria met may be problematic. But there were limitations to the previous study: only one disorder was analyzed, and the data were gathered from the general population using a poorly-validated, fully-structured assessment instrument. Since these findings have widespread implications for the measurement and diagnosis of many mental disorders, they warrant further analysis to determine if the results are robust (a) across different disorders, (b) in a clinical sample, and (c) when using a well-validated assessment instrument. The purpose of the current study, therefore, was to analyze the effects of this measurement problem in other disorders, while addressing the limitations of the previous study.

To accomplish this goal, I analyzed clinical data from all 10 *DSM-IV-TR* PDs that had been gathered through the administration of a well-validated, semi-structured interview of personality pathology (the Structured Interview for DSM-IV Personality; SIDP-IV; Pfohl, Blum, & Zimmerman, 1997). Given the previous literature on the functioning of PD diagnostic criteria, I expected two main findings: 1) that different combinations of diagnostic criteria at the same raw score would result in varying estimations (or "bands") of the latent PD traits, and 2) that the bands would overlap (i.e., similar latent estimates could be reached by combinations with different numbers of endorsed criteria). To test these hypotheses, IRT analyses were used to calculate the latent PD trait estimate associated with all possible response patterns within each PD criteria set. As IRT is well-equipped to consider the functioning of individual criteria when estimating the latent trait, it was considered to be an apt measurement model to address these hypotheses.

2. METHOD

2.1 *Participants*

Participants were 2,165 outpatients who sought treatment from a university-affiliated clinic in the northeastern United States over the past 12 years and who were part of the larger Methods to Improve Diagnostic Assessment and Services (MIDAS; PI: Mark Zimmerman, M.D.) database. The majority of the participants in this sample were female (61%, *n* = 1,323) and Caucasian (87%, *n* = 1,887). Other ethnicities represented were African American (4%, *n* = 97), Portuguese (4%, n = 76), Hispanic (3%, *n* = 59), Asian (1%, *n* = 21) and "other" (1%, *n* = 25). The mean age was 38.5 years (*SD* = 12.9). Most individuals had never been hospitalized (76%, *n* = 1,648), but many reported that a psychiatric illness had interfered with employment at some time (45%, *n* = 964).

Although the main sample consisted of 2,165 participants, additional participants from the MIDAS database were assessed for antisocial and borderline PD (roughly 500 and 700 more, respectively). Data from these additional participants were included in the analyses for these two PDs. The numbers of participants that were included in the analyses for each PD are displayed in Table 2. In terms of gender, ethnicity and age, the demographic percentages between the two groups (those with the full PD data and those only assessed for borderline and/or antisocial PD) were virtually identical. Informed consent was obtained in writing from all participants.

2.2 *Measures*

In addition to measures of Axis I psychopathology, participants were assessed for personality pathology using the SIDP-IV (Pfohl et al., 1997). The SIDP-IV is a semi-structured interview that assesses 13 PDs, but only the 10 PDs officially included in the *DSM-IV* were analyzed in the current study. Each SIDP-IV question was derived from a *DSM-IV* PD diagnostic criterion, and responses are rated on a scale from 0 (criterion is "not present") to 3 (criterion is "strongly present"). The SIDP-IV has been used extensively in previous research, and demonstrates good inter-rater reliability (Jane, Pagan, Turkheimer, Fiedler, & Oltmanns, 2007), even when given by individuals with little assessment background (Balsis, Eaton, Zona, & Oltmanns, 2006).

In the current sample, diagnostic interviews were administered by thoroughly trained individuals with ongoing collaborative supervision. Joint diagnostic agreement from a subset of 29 individuals ranged from good (kappa = 0.61 for cluster B PDs) to excellent (kappa = 1.00 for cluster A PDs), with a mean cluster kappa of 0.77 (Zimmerman, Rothschild, & Chelminski, 2005). And when the reliability of dimensional (rather than categorical) scores were estimated, intraclass correlation coefficients (which ranged from 0.89 to 0.96, $M = 0.93$) for the 10 *DSM-IV* PDs indicated a high level of agreement between interviewers (Zimmerman et al., 2005).

2.3 *Data Analyses*

First, the four point (0-3) responses to each SIDP-IV item were dichotomously recoded so that each criterion could be considered as being either absent or present. Specifically, responses of either 0 ("not present") or 1 ("subthreshold") were recoded to

Table 2

Sample Sizes and Confirmatory Factor Analysis Results for the 10 DSM-IV-TR Personality Disorders

	n	TLI	CFI	RMSEA
Cluster A				
Paranoid	2,152	1.00	1.00	0.01
Schizoid	2,150	0.99	0.99	0.01
Schizotypal	2,136	0.89	0.91	0.04
Cluster B				
Antisocial	2,674	0.98	0.98	0.04
Borderline	2,871	0.98	0.98	0.04
Histrionic	2,147	0.97	0.98	0.02
Narcissistic	2,145	0.98	0.98	0.02
Cluster C				
Avoidant	2,151	1.00	1.00	0.00
Dependent	2,147	0.93	0.93	0.05
Obs-Comp	2,144	0.93	0.94	0.04

Note. TLI = Tucker-Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Squared Error of Approximation; Obs-Comp = Obsessive-Compulsive.

0's, and responses of either 2 ("present") or 3 ("strongly present") were recoded to 1's. This dichotomization allowed for the results to be interpreted in terms of the *DSM-IV-TR* diagnostic process, as the summation of dichotomized item scores were then equivalent to the number of criteria met.

Next, factor analyses and IRT analyses were conducted. IRT analyses rest in part on the assumption that all items on a scale are locally independent—that the specified dimensions account for all significant variability among items. In other words, items' residual variances are not correlated after an IRT model is fit to the data. In unidimensional IRT analyses (such as those utilized in the current study), the local independence assumption can be tested by the fit of a one-factor solution. If the data are unidimensional, then the items can be considered to be locally independent (see Hambleton, Swaminathan, & Rogers, 1991). To test for unidimensionality (and hence local independence), categorical confirmatory factor analyses (CFAs) of the dichotomized data were conducted in Mplus Version 5.2 (Muthén & Muthén, 2007) using the WLSMV (weighted least squares mean and variance adjusted) estimator for each of the 10 *DSM-IV-TR* PDs. Categorical CFA assumes that a continuous latent response variate function underlies the observed categorical responses to each item and uses a threshold parameter to determine the point(s) along the latent dimension where the observed responses are differentiated (i.e., uses tetrachoric or polychoric correlation coefficients; Wirth & Edwards, 2007), making it a superior method over traditional linear factor analysis when analyzing categorical data.

To analyze the latent structure of the 10 *DSM-IV-TR* PDs, I assessed the fit of a one-factor model for each disorder. As outlined by Hu & Bentler (1999), when the following fit indices are met—the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973) and Comparative Fit Index (CFI; Bentler, 1990) values are "close to 0.95," and the root mean squared error of approximation (RMSEA; Steiger & Lind, 1980) is less than 0.06—one can conclude that a hypothesized model provides a relatively good fit to the observed data (p. 1). Since I was assessing the fit of a one-factor model for each of the 10 PD constructs, I considered the disorders whose fit statistics met these recommended cutoffs to be unidimensional.

The results of the 10 CFAs are displayed in Table 2. Using the above criteria as indicators of unidimensionality, seven of the 10 PDs (paranoid, schizoid, antisocial, borderline, histrionic, narcissistic, and avoidant) were immediately considered unidimensional, as they met or exceeded each fit index. Two other PDs, obsessive-compulsive and dependent, met the RMSEA cutoff, but had TLI and CFI values slightly below 0.95. But because their values were still close to 0.95 (either 0.93 or 0.94) and were above 0.90, these data were also considered to be unidimensional (IRT analyses do not require perfect unidimensionality, but rather that there is a dominant factor that accounts for item responses; e.g., Hambleton et al., 1991). Schizotypal PD failed to meet the recommended cutoffs for the unidimensionality assumption, a finding that is consistent with previous literature (e.g., Chmielewski & Watson, 2008; Fossati et al., 2001; Raine et al., 1994). As such, analyses were not carried out for this disorder in the current study. The nine PDs that met the unidimensionality assumption, however, were

considered appropriate for subsequent IRT analysis (as unidimensionality also indicates that the items are locally independent; see Hambleton et al., 1991).[3]

Item parameters were then estimated using two-parameter logistic (2PL) IRT analyses, and were conducted separately for each of the nine unidimensional PDs in MULTILOG Version 7 (Thissen, Chen, & Bock, 2003). The 2PL models provided significantly better fit to the observed data than the more constrained one-parameter logistic models (determined using -2 log likelihood difference (nested chi-square) tests; *df*s = 7-9, all *p*s < .01). And given the current method of data collection (through the administration of a semi-structured interview), there was no theoretical reason to use the three-parameter logistic model (which models a lower asymptote to account for guessing). In all analyses, default settings were used, with these exceptions: the number of cycles allowed in the parameter estimation step was increased to 100 and the number of quadrature points was set to 17 (0.5 standard deviation increments from -4.0 to 4.0). All models converged in less than 45 cycles.

In the 2PL model, the *a* ("discrimination" or "slope") parameter indicates the degree to which an item is related to the underlying construct (here the latent continuum

[3] Latent structure analyses using PD criteria and constructs are many, and the results of these studies (i.e., whether or not the diagnostic criteria within each category are best modeled by a uni- or multidimensional latent structure) are often as varied as their methodologies. One could likely find a study that implicated a multidimensional structure for most every PD, in addition to studies supporting their unidimensionality. In short, consensus regarding the latent structure of many PDs has not yet been reached. The purpose of factor analyses in the current study was not to argue either for or against the findings from previous studies on the latent structure of these constructs, or to make claims regarding the validity of each PD construct; rather, the purpose was to test whether the data were suitable for IRT analyses.

of each PD trait) at a 50% probability; the *b* ("difficulty") parameter indicates the level of the latent trait where the probability of a positive response to the item equals 50%. Together, the *a* and *b* parameters define an item's item characteristic curve (ICC), a function that describes the probability that an item is positively endorsed across the defined levels of the latent trait. (The probability that an item is not endorsed, on the other hand, is simply one minus the probability that it is endorsed. I will term this function an "inverse ICC.") When plotting ICCs, the horizontal axis represents the continuum of the latent PD construct, theta, in standard deviation (*SD*) units; and the vertical axis, which represents the probability that an item will be endorsed, ranges from 0 to 1.

Using these ICCs for each PD, an estimated value score (EVS; Lazarsfeld, 1950) was then calculated for all combinations, or "response patterns," to the PD criteria for each disorder. A response pattern is a numerical representation of the responses to a given set of criteria. The response pattern "1110011," for example, represents a combination with a raw score of five (i.e., five criteria were endorsed); and conveys that criteria one, two, three, six, and seven were endorsed, and that criteria four and five were not.

An EVS is an estimation of the latent trait for a given response pattern and is essentially the mean of a maximum likelihood (ML) line associated with that response pattern. The ML function is obtained by multiplying together the ICCs and inverse ICCs for a particular response pattern, and represents the joint probability that the response pattern would be endorsed at all levels of the latent dimension (see Embretson & Reise,

2000 or Thissen & Orlando, 2001 for excellent reviews of these concepts). An EVS is but one way to estimate the latent trait associated with a response pattern and was used because it takes the full distribution of the ML line into account (contrary to the maximum value of the ML line, for example) without making prior assumptions about the distributions of the latent PD traits (contrary to posterior scoring methods, for example). EVSs are essentially estimated a posteriori (EAP) estimates without a prior, and both scoring methods yield similar patterns of results.

2.4 *The Latent Trait*

Before proceeding to the results, a brief discussion on the description and interpretation of the latent trait, theta, is warranted. Traditionally, theta is thought as a metric of ability or proficiency. When using clinical, personality, or attitudinal data, however, theta may be better interpreted as a metric of the latent trait that the items in a given test were designed to measure. In the case of narcissistic PD, for example, theta is best conceptualized as a dimensional metric of narcissism (as operationalized by its nine diagnostic criteria) that underlies or influences an individual's observed item responses. Likewise, the latent traits underlying the constructs of the avoidant and dependent PDs could be respectively interpreted as metrics of pathological "avoidance" and "dependence."

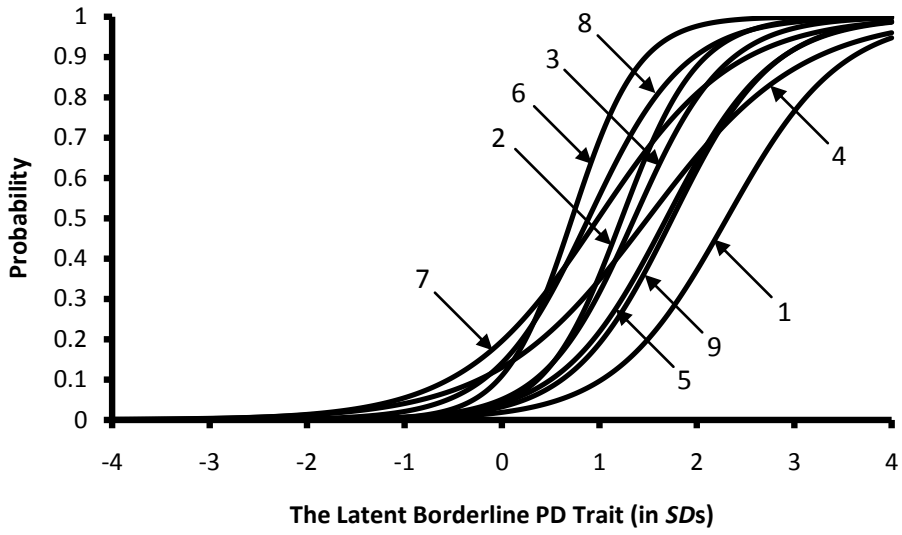
I must note, however, that theta is not necessarily a direct measure of distress or impairment. Though these constructs are surely related (dysfunction is essentially "built in" to the content of the diagnostic criteria that measure the trait; Spitzer & Wakefield, 1999), the most difficult or discriminating criterion (as described by its ICC) may not be

a disorder's most impairing feature. Thus, I suggest that theta be interpreted as a dimensional metric of each PD, until future research can more clearly and comprehensively elucidate its relationship to external measures of distress or impairment. A precise interpretation of theta when using clinical data remains an important area of inquiry. Interested readers should see Reise & Waller (2009) for a more thorough discussion.

3. RESULTS

Because the analyses were similar for each disorder, a detailed explanation of the analyses for one PD (borderline) should clarify the results for the remaining eight PDs. Figure 1 displays the item parameters and ICCs for the nine borderline PD diagnostic criteria. Each ICC represents the probability that a criterion would be endorsed at each level of the latent dimension. As can be seen in Figure 1, each criterion functions (i.e., measures the latent dimension) differently. An individual with 1.0 *SD*s of the borderline trait, for example, would probably not endorse criterion one (roughly a 10% chance), but would most likely endorse criterion six (roughly a 70% chance). Of important note is that the ICCs differ in their discrimination (slope) parameters—steeper-sloping ICCs representing the criteria that are more related to (i.e., better at measuring) the latent trait.

The ICCs in Figure 1, however, only represent the probability that an item would be endorsed. To calculate an EVS for a given response pattern, the ICCs must be reconfigured to reflect which criteria are and are not endorsed. As mentioned previously, the probability that an item would not be endorsed is simply one minus the probability that it would be endorsed. So to represent the item response functions for the response pattern, "100110101," for example, the ICCs for items two, three, six, and eight would have to be "inverted," as seen in the top panel of Figure 2. With the ICCs in this configuration, the next step would be to multiply together the nine ICCs at all levels of the latent dimension to form the ML line, as seen in the bottom panel of Figure 2.



	a	*b*
1	1.71	2.31
2	2.55	1.23
3	2.15	1.36
4	1.27	1.50
5	1.85	1.69
6	2.91	0.72
7	1.43	0.99
8	2.03	0.89
9	1.94	1.75

Figure 1. Item characteristic curves and item parameters for the borderline personality disorder criteria. 1 = Frantic efforts to avoid abandonment; 2 = Intense and unstable relationships; 3 = Identity disturbance; 4 = Self-damaging impulsivity; 5 = Recurrent suicidal activity; 6 = Affective instability; 7 = Feelings of emptiness; 8 = Inappropriate or uncontrollable anger; 9 = Transient, stress-related paranoia or dissociation; *a* = "discrimination" parameter; b = "difficulty" parameter.

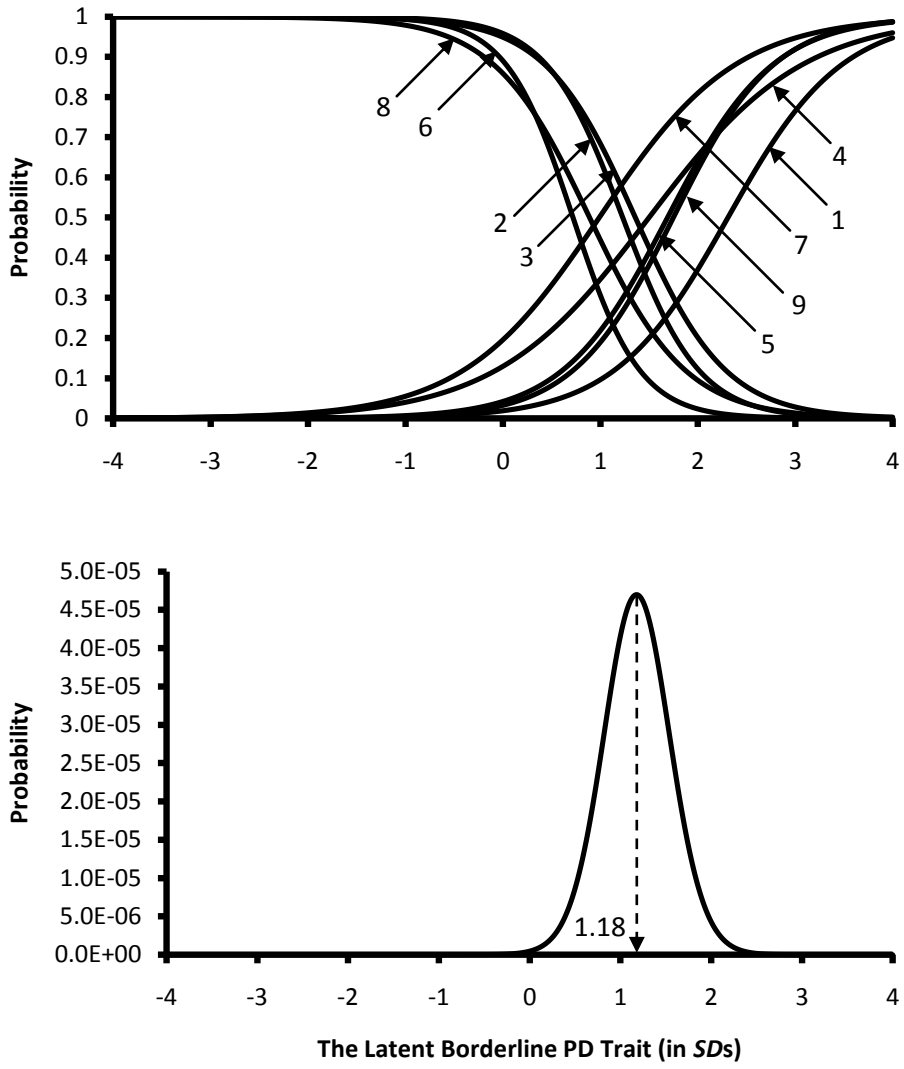


Figure 2. Item characteristic curves for response pattern, "100110101," and its associated maximum likelihood function.

The ML line represents the joint probability of endorsing an observed response pattern across the specified levels of the latent dimension (here, between -4.0 and 4.0 *SD*s). Identifying the level of theta where the ML line peaks is one way of scoring a response pattern (known as the maximum likelihood estimate, or MLE). An EVS, however, is calculated by estimating the average area under the ML line (see Thissen & Orlando, 2001). These two estimation procedures return virtually identical values when the arc of the ML line is roughly symmetrical. For this response pattern (100110101), the MLE (as shown in the bottom panel of Figure 2 for illustrative purposes) and EVS were equal, at 1.18 *SD*s of the latent borderline trait; but this will not always be the case.

The preceding paragraphs outlined the EVS calculations for one borderline PD response pattern. In total, 512 EVSs were calculated for borderline PD—one for each way to respond to its nine diagnostic criteria (see Footnote 1 and Table 1). Each EVS was then plotted along theta at its corresponding raw score, as displayed in Figure 3 (individual data points are difficult to differentiate at certain raw scores). For example, the response pattern from the running example (100110101) is the leftmost (minimum) data point at a raw score of five, with an EVS of 1.18 *SD*s. The response pattern, "011001011," conversely, is the rightmost (maximum) data point among the combinations of five borderline criteria, and had an EVS of 1.69 *SD*s.

At each raw score, response patterns including the most or least discriminating criteria (as determined by their *a* parameters) yield the highest and lowest trait estimates, respectively. Thus, at raw score of five borderline criteria, the response pattern that includes the most-discriminating diagnostic criteria (numbers 2, 3, 6, 8, and 9; see Figure

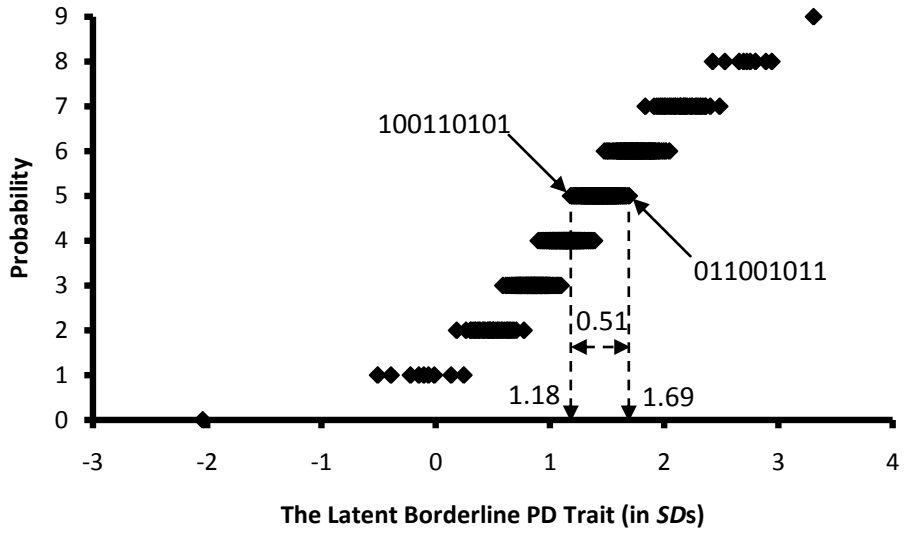


Figure 3. Bands of the borderline personality disorder trait represented at each raw score. A raw score equals the number of diagnostic criteria endorsed; each diamond represents the estimated value score (EVS) for one of the 512 response patterns to the nine borderline diagnostic criteria.

1) will have a greater EVS than that of any other combination of five criteria. Conversely, the response pattern consisting of the five least-discriminating criteria (numbers 1, 4, 5, 7, and 9) will have the lowest trait estimate among all combinations with a raw score of five. And because an item's discrimination parameters conveys how related an item is to its latent dimension (i.e., how well it can measure the construct), the combinations consisting of the criteria most and least related to the construct will respectively represent the maximum and minimum EVS at each raw score.

The difference between the minimum and maximum EVS at each raw score yields that raw score's "band width." At the raw score of five borderline criteria, for example, the band width was calculated by subtracting the minimum EVS (1.18 *SD*s) from the maximum EVS (1.69 *SD*s). Hence, the band width at a raw score of five borderline PD criteria was 0.51 *SD*s (see Figure 3). Band widths for the remaining borderline PD raw scores were calculated in the same manner. Of course, there is only one way to endorse no criteria or all of the criteria, so there are no bands for raw scores consisting of zero and *n* endorsed criteria (where *n* is equal to the total number of diagnostic criteria for a given PD). But for the raw scores with more than one possible response pattern (here, one through eight), the band widths ranged from 0.49 to 0.75 *SD*s of the borderline PD trait, and had a mean width of 0.57 *SD*s. This finding confirmed the first hypothesis (that each raw score would correspond to a range of the latent PD trait).

But the band widths are not the only important characteristic displayed in Figure 3. Of further note is that the bands from one raw score overlap with those from abutting

raw scores. Recall that the minimum EVS at a raw score of five was from response pattern, "100110101," and indicated an estimated 1.18 *SD*s of the latent borderline trait. As can be seen by the EVSs at a raw score of four, many response patterns met or exceeded this estimate. Specifically, 39% (49) of the 126 subthreshold combinations of four criteria had EVSs that was greater than or equal to this minimum at-threshold EVS. In total, the majority (63%, 319) of the 512 borderline EVSs overlapped with those from an abutting raw score. This finding confirmed the second hypothesis (that the bands would overlap).

Similar analyses were then conducted for the remaining PDs (excluding schizotypal). Like the results reported for borderline PD, the following characteristics were of interest for each disorder: The percent of response patterns with EVSs that overlapped with those from other raw scores; the number of subthreshold EVSs that overlapped with an at-threshold EVSs; and the minimum, maximum, and mean band widths. The numerical results of these analyses are displayed in Table 3 (the previously reported values for borderline PD are presented here again for comparative purposes), and the graphical representations of the band widths and overlap for the remaining eight PDs are displayed in Figure 4 (see Figure 3 for the borderline PD bands). The horizontal whisker plots in Figure 4 represent the trait band widths at each raw score; the diamonds represent the mean EVS of each raw score. Item parameters and ICCs for the remaining analyzed PD criteria are displayed in the Appendix.

As can be seen in Figures 3 and 4, the number of criteria met (plotted along the vertical axis) corresponds to bands of the latent PD trait estimates (plotted along the

Table 3

Personality Disorder Trait Raw Score Overlap and Band Width Characteristics

	Overlap	±2	Sub-Diagnostic	Min. Band Width*	Max. Band Width*	Mean Band Width*
Cluster A						
Paranoid	47% (61)	0	13	0.40	0.78	0.58
Schizoid	52% (66)	0	19	0.17	0.99	0.67
Cluster B						
Antisocial	23% (30)	0	1	0.36	0.52	0.42
Borderline	63% (319)	0	39	0.49	0.75	0.57
Histrionic	85% (218)	8	41	0.31	1.24	0.76
Narcissistic	87% (446)	5	71	0.33	0.91	0.58
Cluster C						
Avoidant	49% (63)	0	11	0.42	0.72	0.52
Dependent	74% (189)	0	32	0.32	0.82	0.62
Obs-Comp	94% (242)	23	43	0.46	1.62	1.26

Note. * in *SD* units. Overlap = The total number of response patterns with an expected value score (EVS) that overlapped with an EVS from another raw score; ±2 = The number of EVSs that overlapped with an EVS from a combination of either two less or two more endorsed criteria; Sub-Diagnostic = The total number of subthreshold response patterns with an EVS that overlapped with a diagnostic (at-threshold) EVS.

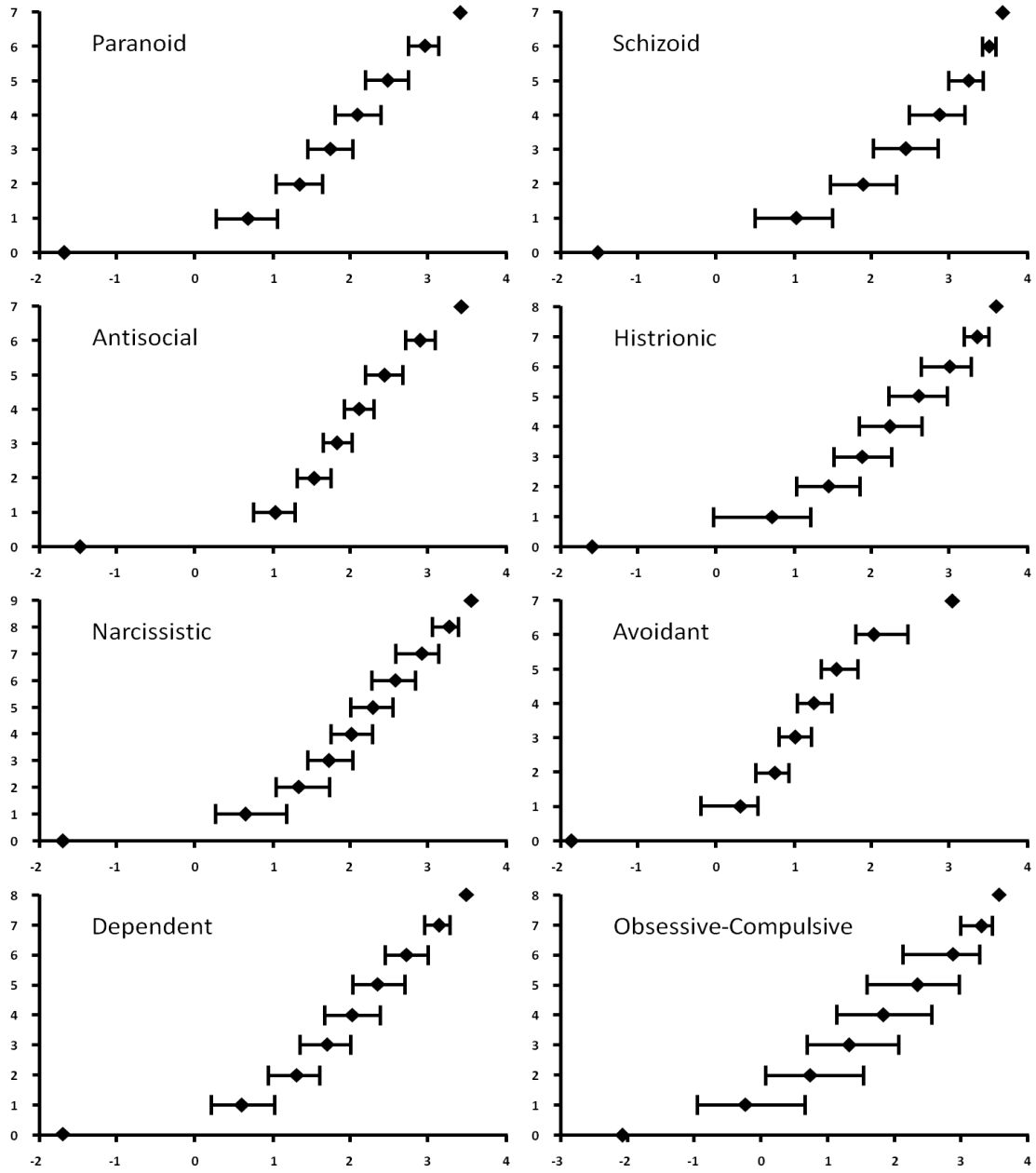


Figure 4. Bands of latent personality disorder traits. In each graph the horizontal axis represents each latent PD trait in *SD* units and the vertical axis represents the raw score (i.e., the number of diagnostic criteria endorsed). Diamonds depict the mean expected value score (EVS) at each raw score.

horizontal axis) that overlap with the trait estimations from abutting raw scores. The knowledge that than an individual meets four borderline PD criteria, for example, conveys only that he or she has somewhere between 0.90 and 1.39 *SD*s of the latent borderline trait (depending on which one of the 126 response patterns of four criteria he or she endorsed). Furthermore, of the 126 ways to endorse four borderline criteria, 96 of them have EVSs that overlap with EVSs from raw scores three and five. In other words, a raw score of four borderline criteria conveys only that an individual falls somewhere along a band of the latent trait almost half of a *SD* wide, and that he or she could have the same estimated trait level as individuals meeting three or five criteria. There were even a few instances of band overlap with the raw scores from combinations with two greater or two fewer endorsed criteria (see the "±2" column in Table 3).

The most problematic instances of overlap, however, occurred around the raw scores that correspond to a disorder's diagnostic threshold. Since criteria combinations at one raw score often overlap with combinations at other raw scores, the possibility arises that subthreshold response patterns could exhibit as much of the latent trait as combinations at the diagnostic threshold (Cooper & Balsis, 2009). In such circumstances, an individual could not technically be diagnosed with a disorder (since he or she did not endorse enough criteria to meet the diagnostic threshold), even though he or she exhibits as much of the latent trait as individual as the diagnostic level. As the "Sub-Diagnostic" column in Table 3 displays, there were 270 subthreshold PD response patterns that exhibited as much or more of the corresponding latent trait as diagnostic (at-threshold or greater) ones. There were even a few (eight) response patterns that were

two criteria below a disorder's diagnostic threshold, yet had EVSs equal to an at-threshold EVS.

In total, 71% (1,634) of the 2,304 response patterns to the nine analyzed PDs had EVSs that overlapped with the trait estimates from another raw score. Put another way, for each disorder the number of PD criteria met only corresponded to a level of the latent trait unique to its own raw score about one quarter (29%) of the time. In all other instances, the trait estimates at one raw score could also be attained by a response pattern consisting of a different number of endorsed criteria.

4. DISCUSSION AND CONCLUSION

Although an abundance of research suggests that the diagnostic criteria for each PD have different psychometric properties, the diagnostic algorithm used in the *DSM-IV-TR* weights each PD criterion equally. In an effort to analyze the consequences of this discrepancy, the current study sought to compare the criteria-counting procedure used in the *DSM* (a method that considers only the number of criteria met) to a model that considers both the number of criteria met and the functional characteristics of a disorder's diagnostic criteria (IRT parameter estimation and response pattern scoring). I examined the ability of these two models to accurately measure the trait dimensions for each of the nine analyzed PDs.

The results from each model comparison could have resulted in one of three outcomes. First, the IRT analyses could have indicated that all criteria for a given PD equally measured the construct. In the context of the current analyses, this would mean that the ICCs for every diagnostic criterion within a PD would have identical discrimination parameters. If each criterion were equally related to the construct, the number of criteria met would be a sufficient statistic for measuring the latent trait. Because every combination with the same number of endorsed criteria would result in the same ML line (and hence the same EVS), there would be no bands: all raw score combinations would correspond to the same latent trait estimate. A simple criterion count could be implemented in this situation.

The second possible result of these analyses could have indicated that each criterion was roughly equivalent in its ability to measure the latent construct. This would mean that the discrimination parameters for a disorder's ICCs would vary only slightly. In the event of this outcome, the number of criteria met would no longer be a sufficient statistic for the estimation of an individual's latent trait level (as there would be narrow bands that formed at each raw score), but the bands could be narrow enough so that they did not overlap with the bands formed by abutting raw scores. If this were the case, using the number of criteria met as a metric for the latent trait would result in the loss of some information (as a criterion count would disregard the varying latent trait estimations at each raw score). But since the bands would not overlap, each response pattern would at least correspond to a "slice" of the latent dimension that was unique to its own raw score.

In the third potential outcome of these model comparisons, the diagnostic criteria for a given disorder would function so differently that the bands formed at each raw score would overlap. In this circumstance, using the number of criteria met as a metric for the latent trait is especially problematic; here, a criterion count does not correspond to a singular point for all combinations of the same raw score (as it did in the first hypothetical outcome), nor does it correspond to a band of the latent trait unique to its own raw score (as it did in the second hypothetical outcome). Rather, the number of criteria met can correspond to latent trait estimates (e.g., EVSs) from combinations with a different number of endorsed criteria. Hence, the number of criteria met would be an imprecise metric for PD trait estimation in this situation.

The current results indicated that each of the nine analyzed PDs fell into this third category. Not only did raw scores correspond to bands of each PD trait dimension, but the bands formed at one raw score also overlapped with the bands from abutting raw scores (see Figures 3 and 4). These results suggest that using a criteria count to measure and diagnose PDs yields only a coarse estimation of an individual's level of the latent trait. Hence, it is important to consider the psychometric properties of a disorder's criteria in the diagnostic process. Without considering these psychometric properties, potentially valuable information will be lost—information that could suggest a different diagnostic decision (in the case of certain subthreshold combinations of criteria, for example).

There were also differences in the band widths and raw score overlaps among the analyzed PDs (see Table 3), indicating that criterion counts were more problematic in some disorders than others. For example, antisocial PD had relatively narrow bands (M = 0.42 SDs) and a small number of overlapping EVSs (23%), whereas obsessive-compulsive PD had very wide bands (M = 1.26 SDs) and a large number (94%) of overlapping EVSs. Since these two disorders represent the two extremes of the model comparison discrepancies, I will discuss them in further detail. I will also discuss borderline PD, as it the most researched of the nine analyzed disorders. And since we had previously analyzed the criteria for schizoid PD (using data gathered from the general population with a different assessment instrument; Cooper & Balsis, 2009), I will also discuss how the current schizoid analyses compared to the results of the

previous study. However, I will not discuss in detail the results from the remaining five analyzed PDs.

4.1 *Borderline Personality Disorder*

Borderline is one of the few PDs for which the effects of endorsing combinations of criteria have been analyzed. Consistent with the results of the current study, previous research also shows that the criteria differ in their ability to measure the construct and that certain sub-threshold combinations of criteria can reliably estimate the presence of the disorder. Some have suggested that the presence of four, three, or even two specific criteria may be able to reliably diagnose borderline PD (e.g., Blais, Hilsenroth, & Fowler, 1999; Clifton & Pilkonis, 2007; Hurt et al., 1990; Nurnberg, Hurt, Feldman, & Suh, 1988; Nurnberg et al., 1991; Reich, 1990; Widiger Frances, Warner, & Bluhm, 1986), and others have even suggested that one can "make the diagnosis with relative confidence in certain patients with three, two, or even just one feature!" (Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983, p. 272).

The criteria that were found to be the best markers of the disorder in these previous studies were remarkably consistent. In almost all instances, criteria two (unstable and intense relationships) and six (affective instability) were listed as important diagnostic features; criteria three (identity disturbance) and eight (intense anger) were implicated in about half of the studies. Interestingly, these were the four criteria with the highest discrimination parameters in the current study. The response pattern consisting of these four criteria had highest EVS among all combinations of four borderline criteria and indicated as much or more of the latent trait than 49 combinations

of five criteria. Criteria two and six were the same criteria found to have the best joint probability for successful borderline PD case detection (Hurt et al., 1990) and are also rated as among the most prototypical descriptors of borderline clients (e.g., Burgmer, Jessen, Freyberger, 2000; Shedler & Westen, 2004). The current findings, taken together with this previous research, suggest that these two criteria (two and six) may be especially important to consider in the context of subthreshold combinations, as they are often associated with diagnostic levels of the latent trait.

4.2 *Schizoid Personality Disorder*

One of the purposes of the current study was to replicate the analyses previously conducted for schizoid PD (Cooper & Balsis, 2009). Although the implications were similar in both studies (in terms of band width and raw score overlap), there were some notable differences—especially in the *b* ("difficulty") parameters of the diagnostic criteria. But first, the general comparisons: The mean band widths were slightly wider in the previous analyses (0.89 *SD*s) than in the current analyses (0.67 *SD*s). Relatedly, the number of EVSs that overlapped with the EVS from other raw scores was slightly greater in the previous analyses (78) than in the current analyses (66). The number of subthreshold response patterns that overlapped with a diagnostic response pattern, however, was fewer in the previous analyses (15) than in the current analyses (19). Hence, there were slight differences in between studies, but the pattern of results was similar. Indeed, the rank-order stability of the 128 EVSs between studies was remarkably high (Pearson's $r = 0.98$, $p < .001$), indicating that the estimates between studies were very consistent.

What is not captured in the above comparisons, however, is that the schizoid bands in the current study have shifted a mean distance of 1.18 *SD*s to the right of the previous estimates—a difference due to changes in the criteria's difficulty parameters between analyses. Though the similarly and rank-order stability of the discrimination parameters between studies was high (Pearson's $r = .93$, $p = .002$), the criteria's difficulty parameters often varied significantly. I attribute these differences to the assessment methods and samples used in each study. Data from the previous study were gathered by administering a fully structured interview to participants from the general population, whereas data from the current study were gathered by administering a semi-structured interview to a sample of treatment-seeking outpatients. And since those with schizoid PD are unlikely to seek treatment (e.g., Tyrer, Mitchard, Methuen, & Ranger, 2003), the epidemiological data analyzed in the previous study may have captured a larger amount of individuals with schizoid traits. Furthermore, even though both assessment instruments were based on the same *DSM-IV* criteria, the wording of the interview questions was not identical; and small wording changes can have dramatic effects on endorsement frequencies of the "same" diagnostic criteria (Blashfield, Blum, & Pfohl, 1992). Even with these differences, however, identical conclusions can still be drawn from both the previous and current study: that using the number of PD criteria met to determine an individual's level of the construct is imprecise and that there are subthreshold combinations of criteria that indicate diagnostic levels of the latent trait.

4.3 *Antisocial Personality Disorder*

Antisocial PD was unique among the nine analyzed disorders in that its diagnostic criteria functioned fairly similarly (as defined by their item parameters/ICCs; see the Appendix). Since each item was similar in its ability to measure the latent antisocial trait (i.e., had similar discrimination parameters), the mean band width formed at each raw score was relatively narrow (0.42 *SD*s; see Table 2). And although a substantial number of EVSs did overlap with the EVSs from another raw score (23%), there was only one subthreshold response pattern that resulted in a trait estimate equivalent to an at-threshold EVS. This single pattern, consisting of criterion one (nonconformance to social norms and laws) and seven (lack of remorse for one's actions), evidenced as much of the latent antisocial trait as 10 at-threshold response patterns.

Of the nine analyzed disorders, the criteria for antisocial PD conformed best to the current polythetic diagnostic system: it had the narrowest mean band width, the lowest number of overlapping response patterns, and only one overlapping subthreshold response pattern. It is probably no coincidence, then, that the selection of these criteria was partially informed by their field-tested psychometric properties (see Widiger et al., 1996). Criteria that were only marginally related to the other diagnostic features were considered for deletion or revision. In other words, potential criteria that performed poorly were "weeded out," which made for a fairly homogenous final set of diagnostic criteria. If the poorly-functioning criteria had been retained, I surmise that the band widths and raw score overlap would have been wider.

4.4 *Obsessive-Compulsive Personality Disorder*

In stark contrast to the antisocial PD criteria, most of the obsessive-compulsive PD (OCPD) criteria functioned very differently (see the Appendix). As a result, a raw score was an especially coarse estimation of the latent trait. OCPD was the only analyzed disorder with a mean band width greater than 1.0 *SD*; it also evidenced the highest percentage of overlapping response patterns (94%). EVSs from 43 subthreshold response patterns overlapped with a diagnostic (at-threshold) EVS, and four of these were from response patterns with a raw score two criteria below the diagnostic threshold. In other words, some combinations of two criteria had an estimated trait level that exceeded the EVSs from combinations of four criteria. The two criteria driving the overlap with higher raw scores (i.e., the criteria with the highest discrimination parameters) were numbers one (preoccupation with details and order) and two (perfectionism that interferes with task completion). Interestingly, these two criteria have also consistently emerged in previous research as being among the most related criteria to the construct (Farmer & Chapman, 2002; Grilo, 2004; Hummelen, Wilberg, Pedersen, Karterud, 2008; Sanislow et al., 2002).

The varying discrimination parameters for OCPD, however, suggest that many criteria are only marginally related to the latent dimension. Hence, the construct, as it is currently operationalized by its diagnostic criteria, seems rather diffuse—a conclusion also drawn by other researchers (e.g., Farmer & Chapman, 2002, referred to OCPD as a "collection of weakly related features," p. 296). Like others (e.g., Farmer & Chapman; Hummelen et al., 2008), I conclude that psychometric properties of the current OCPD

criteria are poor. Both the previous and current research on OCPD seems to suggest that the classification and conceptualization of this disorder is in need of revision. Specifically, research is needed to develop criteria that can better measure this construct.

4.5 *Implications for the Classification of Personality Disorders*

There is widespread dissatisfaction with the current classification of PDs (see Bernstein, Iscan, Maser, 2007), and the current analyses highlight additional problems with this system. As evidenced by the substantial amount of raw score overlap, these results suggest that the *DSM-IV-TR* diagnostic algorithm does not adequately measure its PD constructs. Since the number of criteria met corresponds to a band of each latent trait dimension, using raw scores to assess PD is inadequate. And because the bands formed at each raw score overlap with those from abutting raw scores, there can be subthreshold combinations of criteria that correspond to diagnostic levels of PD traits. In short, it seems that the current algorithm for measuring and diagnosing PD is in need of revision.

Not surprisingly, the PDs are poised for great change in the *DSM-V*, and the PD work group is considering many alternative measurement models (see Skodol, Bender, & Oldham, 2009). The most straightforward proposals advocate for the dimensionalization of the current PD constructs (see Widiger & Simonsen, 2005), whereas other proposals suggest that personality pathology would be better modeled using a general dimensional structure of personality (see Skodol et al., 2009 for a brief discussion of a few such models). Results of the current study are more readily applicable to some proposals than others, but the general implications should be relevant

to any potential model of personality pathology: Whatever model is ultimately used in the *DSM-V*, there will likely be items to measure personality dimensions, or diagnostic criteria to measure PD constructs (categories or prototypes). And for their accurate measurement, the contribution of each item or criterion should be considered. Otherwise, the same measurement concerns outlined in the current study (i.e., bands and raw score overlap) could plague subsequent PD classification models.

4.6 *Implications for a Polythetic Diagnostic System*

The measurement concerns raised regarding the PDs are also relevant for the classification and diagnosis of many other polythetically-defined mental disorders. For example, IRT has also been used to analyze the diagnostic criteria for other polythetically-defined disorders such as depression (Aggen et al., 2005), conduct disorder (Gelhorn et al., 2009), panic attack (Ietsugu, Sukigara, & Furukawa, 2007), and substance use disorders (e.g., Langenbucher et al., 2004). A consistent finding in all of these studies is that most diagnostic criteria differed in their ability to measure their disorder construct (the characteristic that causes the bands at each raw score). Hence, the same patterns of results reported in the current study are surely evident in other disorders. And given the possibility that a new model will be implemented for measurement and diagnosis of the *DSM-V* PDs, the implications of the current study may ultimately be most relevant for the disorders that will retain their polythetic diagnostic format. Therefore, a discussion of the general implications for, and the potential improvements to, the current polythetic system is warranted.

From a psychometric perspective, the traditional (unit-weighted) polythetic system is most appropriate only when each criterion is equally able to measure the latent construct. In other words, if a criteria-counting algorithm is utilized in the diagnosis of mental disorders, every criterion in a diagnostic set should be equally related to the disorder construct. In the current study, the consequences of violating this principle are evident in the band widths at each raw score. When certain criteria are more related to a disorder construct than others, accounting for which criteria are endorsed becomes important.

This problem was exemplified by the OCPD criteria. Because the OCPD criteria varied widely in their ability to measure the latent construct, a simple criterion count was inadequate for measuring this disorder. On the other hand, antisocial PD, because its diagnostic criteria were roughly similar in their ability to measure the latent trait, conformed best to the *DSM-IV-TR*'s polythetic (criteria counting) measurement model. Its raw scores, while still yielding instances of overlap, corresponded to relatively narrow bands of the latent dimension. Put simply, as the discrimination parameters of a disorder's criteria become more similar, a raw score approaches a more accurate approximation of the latent trait. Hence, one way of addressing the concerns raised by the current analyses would be to develop criteria that are equally able to measure the latent construct (i.e., criteria that have identical discrimination parameters). Under this condition, counting the number of criteria met would be adequate to assess an individual's level of the disorder trait.

Another possibility would be to weight individual diagnostic criteria based on their ability to measure the disorder construct. Instead of using a raw score, thresholds could be based on a weighted sum of the endorsed criteria. The potential benefits of a weighting scheme have been acknowledged for years (e.g., Widiger et al., 1988), and this solution also seems more consistent with how clinicians use the diagnostic criteria (Davis et al., 1993; Evans et al., 2002; Kim & Ahn, 2002; Morey & Ochoa, 1989). Unfortunately, at present it is not perfectly clear how to optimally weight a disorder's diagnostic criteria. Weighting algorithms have shown promise in other disorders (e.g., alcohol dependence; Langenbucher, Morgenstern, Labouvie, Miller, & Nathan, 1996), but future research is still needed in this area.

A pertinent feature of 2PL IRT models, and one that could help inform a criterion weighting scheme, is that the sum of the discrimination parameters of a response pattern's endorsed criteria yields a sufficient statistic for measuring the latent trait (e.g., de Ayala, 2008). Hence, a criterion's discrimination parameter could possibly serve as its weight, and the weights of the criteria that constitute the minimum at-threshold EVS for a given disorder could then be summed to form a weighted diagnostic threshold.

To illustrate, recall that the minimum at-threshold EVS for borderline PD was from the response pattern, "100110101" (see Figure 3). The sum of the discrimination parameters (displayed in Figure 1) of these five endorsed criteria is 8.2. This number, as the lowest weighted value associated with the raw score diagnostic threshold of five endorsed criteria, could serve as the weighted threshold for borderline PD. The benefit of using a weighted threshold is that the hundreds of previously "subthreshold"

combinations of criteria whose trait estimates overlapped with an at-threshold EVS (see the "Sub-Diagnostic" column in Table 3) could now also be considered for diagnosis. For example, the borderline PD response pattern, "111000001," while consisting of only four endorsed criteria (subthreshold by the *DSM-IV-TR* standard), has a weighted sum of 8.4 (rounded to a single decimal point; see Figure 1). This value exceeds the weighted threshold value of 8.2; and shows that, when using this algorithm, an individual meeting a fewer number of diagnostic criteria could still be eligible for a diagnosis if he or she endorses features that are highly related to the latent construct.

A similar solution would be to simply use IRT scoring techniques, such as the one outlined in the current study. When the item parameters for a disorder's diagnostic criteria are known, an individual's level of the latent disorder trait could be determined by entering into a computer program their response pattern to the criteria (i.e., which criteria were and were not endorsed; Cooper & Balsis, 2009). These procedures offer a more accurate estimation of an individual's level of a disorder trait than can be gained by counting criteria alone, and would be relatively easy to implement in future diagnostic systems.

4.7 *Limitations and Future Directions*

While I calculated EVSs for every response pattern for the nine analyzed PDs, all combinations were not present in the current sample's data. Given large enough samples, however, all criteria combinations will likely occur. For example, all 128 possible ways to endorse the schizoid criteria are present in the National Epidemiological Survey on Alcohol and Related Conditions data (NESARC; over

40,000 respondents; Cooper & Balsis, 2009). Because of this, I decided to calculate EVSs for criteria combinations even if they were not present in the current sample's data. However, an EVS does not return a readily interpretable trait estimate for the response patterns where no criteria are endorsed and criteria are endorsed (Embretson & Reise, 2000; Thissen & Orlando, 2001). Because the ML line increases to positive infinity for these response patterns, their EVSs are dependent upon the chosen range of theta. This limitation could have been circumvented by also multiplying a prior (usually Gaussian) distribution with the ML line, but was not done in the current analyses because the population distributions of PD traits are not known.

Another limitation of the current study was that functional impairment and distress were not directly measured. The diagnosis of a mental disorder is often a multi-step process; and in addition to assessing whether individuals have "enough" of the disorder construct to warrant a diagnosis (which is the aspect of the diagnostic process that the current study addresses), the *DSM* often sets forth the requirement that individuals also experience a clinically significant level of distress or impairment. As discussed in the Method section, latent trait estimates may not directly correspond with distress or impairment; hence, additional research is needed to determine if the subthreshold combinations of criteria identified in the current study also correspond to clinically significant levels of distress. Previous research has indicated that latent trait estimates are good predictors of external criteria (e.g., Aggen et al., 2004), and other studies have found that subthreshold combinations of criteria are often related to significant distress and impairment (e.g., Backenstrass et al., 2006; Blagov, Bradley, &

Westen, 2007; see Maser & Patterson, 2002, p. 874-876; Rucci et al., 2003); but future research is still needed to rigorously analyze the relationship between latent trait estimates and external measures of distress or impairment.

Because I was interested in assessing the *DSM-IV-TR* algorithm for diagnosing PDs, I chose to dichotomize the four-point response scale to the SIDP-IV interview items. Although this allowed for a direct comparison with the current PD diagnostic algorithm, it also likely resulted in the loss of some information. Polytomous IRT models could be used to analyze the full response spectrum to the SIDP-IV items; but these models were not used in the current study, as the results would have related more to the psychometric properties of the SIDP-IV than to the current *DSM* diagnostic process. Incorporating a dimensional component into the assessment of disorder features, however, has advantages (Helzer, Kraemer, & Krueger, 2006), and future research should explore the benefits of including this additional information.

A final concern, and one related to the choice to dichotomize the SIDP-IV data, is that I used the *DSM-IV-TR* PD categories as the basis for the current analyses. When all of the PD criteria are included in a single factor analytic model, the results usually fail to support the current diagnostic categories and clusters (see Sheets & Craighead, 2007). The excessive amount of diagnostic comorbidity among *DSM-IV-TR* PD constructs also suggests that the current classification may not be accurately modeling the "true" structure of personality pathology. Although the factor analyses indicated that the nine analyzed constructs were suitable for IRT analyses (i.e., that their criteria were satisfactorily unidimensional and locally independent), I did not test alternate structural

models (see Footnote 3). Hence, the results of the current study may have differed if I had used another model of personality pathology.

4.8 *Conclusions*

The imprecision of the current diagnostic algorithm for PDs is potentially problematic. Because the *DSM-IV-TR* equally weights diagnostic criteria that indicate different levels of their respective disorders, a raw score (the number of criteria met) does not indicate a singular level of the disorder; rather, it corresponds to a band of its corresponding latent PD trait. For a unit-weighted polythetic diagnostic system to be accurately used in the diagnosis of mental disorders, a disorder's diagnostic criteria should all equivalently measure the construct. If they do not, the potential for raw score overlap arises. This is especially problematic when subthreshold combinations of criteria evidence diagnostic levels of the disorder trait, as they do in the nine analyzed PDs. These results highlight the importance of accounting for the differential contributions of individual criteria (and hence combinations of criteria) in the assessment of polythetically-defined mental disorders, and suggest that the implementation of a weighting scheme may serve as an initial step to improve diagnostic accuracy.

REFERENCES

Aggen, S. H., Neale, M. C., & Kendler, K. S. (2005). DSM criteria for major depression: Evaluating symptom patterns using latent-trait item response models. *Psychological Medicine*, *35*, 475-487.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Backenstrass, M., Frank, A., Joest, K., Hingmann, S., Mundt, C., & Kronmüller, K. (2006). A comparative study of nonspecific depressive symptoms and minor depression regarding functional impairment and associated characteristics in primary care. *Comprehensive Psychiatry*, *47*, 35-41.

Balsis, S., Eaton, N. R., Zona, D. M., & Oltmanns, T. F. (2006). Teaching advanced psychopathology: A method that promotes basic undergraduate clinical and research experience. *Teaching of Psychology*, *33*, 242-245.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bernstein, D. P., Iscan, C., Maser, J. (2007). Opinions of personality disorder experts regarding the *DSM-IV* personality disorders classification system. *Journal of Personality Disorders*, *21*, 536-551.

Bienfait, K. L. (2008). Empirical validation of the DSM-IV's personality disorder criteria: An item response theory analysis. (Doctoral dissertation, St. John's University, 2008). *Dissertation Abstracts International*, *86*, 7081.

Blagov, P. S., Bradley, R., & Westen, D. (2007). Under the Axis II radar: Clinically relevant personality constellations that escape DSM-IV diagnoses. *The Journal of Nervous and Mental Disease*, *195*, 477-483.

Blais, M. A., Hilsenroth, M. J., & Fowler, C. (1999). Diagnostic efficiency and hierarchical functioning of the DSM-IV borderline personality disorder criteria. *The Journal of Nervous and Mental Disease*, *187*, 167-173.

Blashfield, R. K., Blum, N., & Pfohl, B. (1992). The effects of changing Axis II diagnostic criteria. *Comprehensive Psychiatry*, *33*, 245-252.

Burgmer, M., Jessen, F., & Freyberger, H. J. (2000). Polythetic diagnostic approach to the borderline personality disorder: The valency of the single criterion in the concept of professional therapists. *Psychopathology*, *33*, 119-124.

Chmielewski, M., & Watson, D. (2008). The heterogeneous structure of schizotypal personality disorder: Item-level factors of the Schizotypal Personality Questionnaire and their associations with obsessive-compulsive disorder symptoms, dissociative tendencies, and normal personality. *Journal of Abnormal Psychology*, *117*, 364-376.

Clarke, D. M., & McKenzie, D. P. (1994). A caution on the use of cut-points applied to screening instruments or diagnostic criteria. *Journal of Psychiatric Research*, *28*, 185-188.

Clarkin, J. F., Widiger, T. A., Frances, A., Hurt, S. M., & Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. *Journal of Abnormal Psychology*, *92*, 263-275.

Clifton, A., & Pilkonis, P. A. (2007). Evidence for a single latent class of *Diagnostic and Statistical Manual of Mental Disorders* borderline personality pathology. *Comprehensive Psychiatry*, *48*, 70-78.

Cooper, L. D., & Balsis, S. (2009). When less is more: How fewer diagnostic criteria can indicate greater severity. *Psychological Assessment*, *21*, 285-293.

Davis, R. T., Blashfield, R. K., & McElroy, R. A., Jr. (1993). Weighting criteria in the diagnosis of a personality disorder: A demonstration. *Journal of Abnormal Psychology*, *102*, 319-322.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NJ: Guilford Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahway, NJ: Lawrence Erlbaum Associates Publishers.

Evans, D. L., Herbert, J. D., Nelson-Gray, R. O., & Gaudiano, B. A. (2002). Determinants of diagnostic prototypicality judgments of the personality disorders. *Journal of Personality Disorders*, *16*, 95-106.

Farmer, R. F., & Chapman, A. L. (2002). Evaluation of DSM-IV personality disorder criteria as assessed by the Structured Clinical Inventory for DSM-IV Personality Disorders. *Comprehensive Psychiatry*, *43*, 285-300.

Feske, U., Kirisci, L., Tarter, R. E., & Pilkonis, P. A. (2007). An application of item response theory to the DSM-III-R criteria for borderline personality disorder. *Journal of Personality Disorders*, *21*, 418-433.

Fossati, A., Maffei, C., Battaglia, M., Bagnato, M., Donati, D., Donini, M., et al. (2001). Latent class analysis of *DSM-IV* schizotypal personality disorder criteria in psychiatric patients. *Schizophrenia Bulletin*, *27*, 59-71.

Frances, A., Pincus, H. A., Widiger, T. A., Davis, W. W., & First, M. B. (1990). *DSM-IV*: Work in progress. *American Journal of Psychiatry*, *147*, 1439-1448.

Gelhorn, H., Hartman, C., Sakai, J., Mikulich-Gilbertson, S., Stallings, M., Young, S., et al. (2009). An item response theory analysis of *DSM-IV* conduct disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*, 42-50.

Grilo, C. M. (2004). Diagnostic efficiency of DSM-IV criteria for obsessive compulsive personality disorder in patients with binge eating disorder. *Behaviour Research and Therapy*, *42*, 57-65.

Grilo, C. M., McGlashan, T. H., Morey, L. C., Gunderson, J. G., Skodol, A. E., Shea, M. T., et al. (2001). Internal consistency, intercriterion overlap and diagnostic efficiency of criteria sets for DSM-IV schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders. *Acta Psychiatrica Scandinavica*, *104*, 264-272.

Gunderson, J. (1998). DSM-IV personality disorders: Final overview. In T. A. Widiger, A. J. Frances, H. A. Pincus, R. Ross, M. B. First, W. Davis, et al. (Eds.), *DSM-IV sourcebook* (Vol. 4, pp. 1123-1140). Arlington, VA: American Psychiatric Publishing, Inc.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Helzer, J. E., Kraemer, H. C., & Krueger, R. F. (2006). The feasibility and need for dimensional psychiatric diagnoses. *Psychological Medicine, 36*, 1671-1680.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hummelen, B., Wilberg, T., Pedersen, G., Karterud, S. (2008). The quality of the DSM-IV obsessive-compulsive personality disorder construct as a prototype category. *Journal of Nervous and Mental Disease, 196*, 446-455.

Hurt, S. W., Clarkin, J. F., Widiger, T. A., Fyer, M. R., Sullivan, T., Stone, M. H., et al. (1990). Evaluation of DSM-III decision rules for case detection using joint conditional probability structures. *Journal of Personality Disorders, 4*, 121-130.

Ietsugu, T., Sukigara, M., & Furukawa, T. A. (2007). Evaluation of diagnostic criteria for panic attack using item response theory: Findings from the National Comorbidity Study in USA. *Journal of Affective Disorders, 104*, 197-201.

Jane, J. S., Pagan, J. L., Turkheimer, E., Fielder, E. R., & Oltmanns, T. F. (2006). The interrater reliability of the Structured Interview for DSM-IV Personality. *Comprehensive Psychiatry, 47*, 368-375.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General, 131*, 451-476.

Langenbucher, J. W., Morgenstern, J., Labouvie, E., Miller, K. J., & Nathan, P. E. (1996). On criterion weighting in the *DSM-IV*. *Journal of Consulting and Clinical Psychology*, *64*, 343-356.

Langenbucher, J. W., Labouvie, E., Martin, C. S., Sanjuan, P. M., Bavly, L., Kirisci, L. et al. (2004). An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in *DSM-IV*. *Journal of Abnormal Psychology*, *113*, 72-80.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarseld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 362-412). New York: Wiley.

Livesley, W. J., Reiffer, L. I., Sheldon, A. E. R., & West, M. (1987). Prototypicality ratings of DSM-III criteria for personality disorders. *The Journal of Nervous and Mental Disease*, *175*, 395-401.

Maser, J. D., & Patterson, T. (2002). Spectrum and nosology: Implications for DSM-V. *Psychiatric Clinics of North America*, *25*, 855-885.

Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education*, *9*, 3-8.

Morey, L. C., & Ochoa, E. S. (1989). An investigation of adherence to diagnostic criteria: Clinical diagnosis of the DSM-III personality disorders. *Journal of Personality Disorders*, *3*, 180-192.

Muthén, L. K., & Muthén, B. O. (2007). Mplus (Version 5.2) [Computer software]. Los Angeles, CA: Muthén & Muthén.

Nurnberg, H. G., Hurt, S. W., Feldman, A., & Suh, R. (1988). Evaluation of diagnostic criteria for borderline personality disorder. *American Journal of Psychiatry*, *145*, 1280-1284.

Nurnberg, H. G., Raskin, M., Levine, P. E., Pollack, S., Siegel, O., & Prince, R. (1991). Hierarchy of DSM-III-R criteria efficiency for the diagnosis of borderline personality disorder. *Journal of Personality Disorders*, *5*, 211-224.

Pfohl, B., Coryell, W., Zimmerman, M., & Stangl, D. (1986). DSM-III personality disorders: Overlap and internal consistency of individual DSM-III criteria. *Comprehensive Psychiatry*, *27*, 21-34.

Pfohl, B., Blum, N., & Zimmerman, M. (1997). *Structured Interview for DSM-IV Personality*. Washington, DC: American Psychiatric Press.

Raine, A., Reynolds, C., Lencz, T., Scerbo, A., Triphon, N., & Kim, D. (1994). Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophrenia Bulletin*, *20*, 191-201.

Reich, J. (1990). Criteria for diagnosing DSM-III borderline personality disorder. *Annals of Clinical Psychiatry*, *2*, 189-197.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27-48.

Rucci, P., Gheradi, S., Tansella, M., Piccinelli, M., Berardi, D., Bisoffi, G. et al. (2003). Subthreshold psychiatric disorders in primary care: Prevalence and associated characteristics. *Journal of Affective Disorders*, *76*, 171-181.

Sakashita, C., Slade, T., & Andrews, G. (2007). Empirical investigation of two assumptions in the diagnosis of DSM-IV major depressive episode. *Australian and New Zealand Journal of Psychiatry*, *41*, 17-23.

Sanislow, C. A., Morey, L. C., Grilo, C. M., Gunderson, J. G., Shea, M. T., Skodol, A. E., et al. (2002). Confirmatory factor analysis of DSM-IV borderline, schizotypal, avoidant, and obsessive-compulsive personality disorders: Findings from the Collaborative Longitudinal Personality Disorders Study. *Acta Psychiatrica Scandinavica*, *105*, 28-36.

Shedler, J., & Westen, D. (2004). Refining personality disorder diagnosis: Integrating science and practice. *American Journal of Psychiatry*, *161*, 1350-1365.

Sheets, E., & Craighead, W. E. (2007). Toward an empirically based classification of personality pathology. *Clinical Psychology: Science and Practice*, *14*, 77-93.

Skodol, A. E., Bender, D. S., & Oldham, J. M. (2009). Future directions: Toward DSM-V. In J. M. Oldham, A. E. Skodol, & D. S. Bender (Eds.), *Essentials of personality disorders* (pp. 381-392). Arlington, VA: American Psychiatric Publishing, Inc.

Spitzer, R. L., & Wakefield, J. C. (1999). DSM-IV diagnostic criterion for clinical significance: Does it help solve the false positives problem? *American Journal of Psychiatry*, *156*, 1856-1864.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainter (Eds.), *Test Scoring* (pp. 73-140). Mahwah, NJ: Lawrence Erlabaum.

Thissen, D., Chen, W-H, & Bock, R. D. (2003). Multilog (version 7) [Computer software]. Lincolnwood, IN: Scientific Software, Inc.

Tyrer, P., Mitchard, S., Methuen, C., & Ranger, M. (2003). Treatment rejecting and treatment seeking personality disorders: Type R and Type S. *Journal of Personality Disorders*, *17*, 263-268.

Widiger, T. A., Cadoret, R., Hare, R., Robins, L., Rutherford, M., Zanarini, M., et al. (1996). *DSM-IV* antisocial personality disorder field trial. *Journal of Abnormal Psychology*, *105*, 3-16.

Widiger, T. A., Frances, A., Warner, L., & Bluhm, C. (1986). Diagnostic criteria for borderline and schizotypal personality disorders. *Journal of Abnormal Psychology*, *95*, 43-51.

Widiger, T. A., Frances, A., Spitzer, R. L., & Williams, J. B. W. (1988). The DSM-III-R personality disorders: An overview. *American Journal of Psychiatry*, *145*, 786-795.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and *DSM-III*. *Archives of General Psychiatry*, *41*, 1005-1012.

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders*, *19*, 110-130.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79.

Zimmerman, M., Rothschild, L., & Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *American Journal of Psychiatry*, *162*, 1911-1918.

APPENDIX

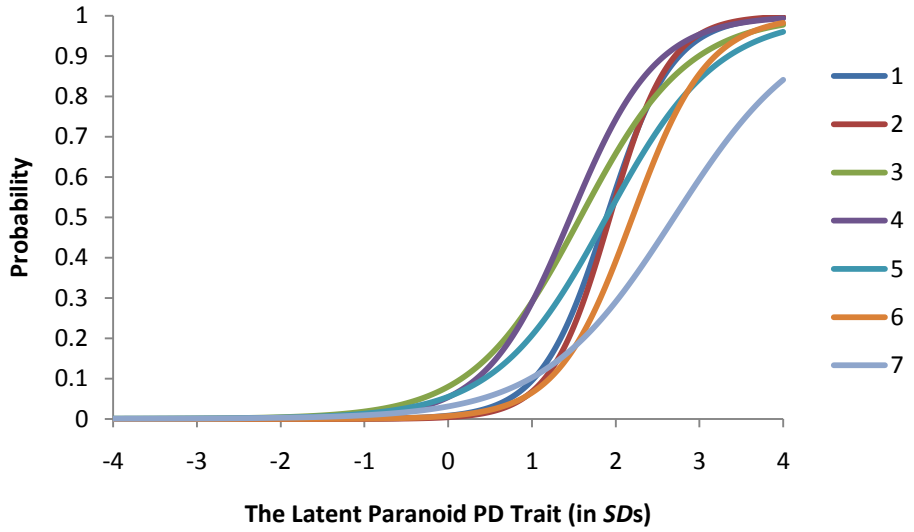
Paranoid Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Is suspecting of others*	2.53	1.89
2	Is distrusting of others*	2.82	1.93
3	Is reluctant to confide in others*	1.55	1.58
4	Reads hidden meanings into things	1.96	1.46
5	Persistently bears grudges	1.51	1.89
6	Perceives attacks to self*	2.22	2.20
7	Doubts fidelity of significant other*	1.28	2.70

Note. * = without sufficient basis, justification, etc.; *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



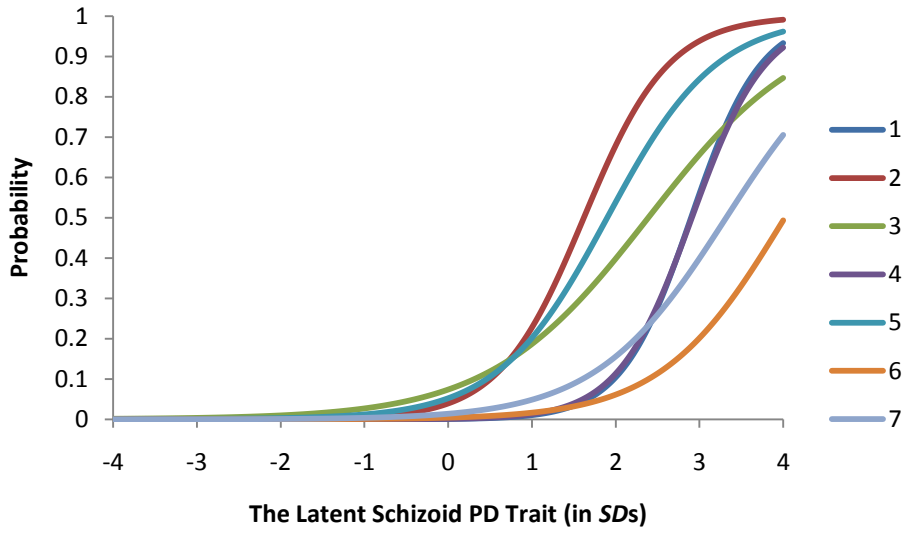
Schizoid Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Does not desire close relationships	2.39	2.90
2	Usually chooses solitary activities	1.97	1.62
3	Has little interest in sexual activities	1.06	2.39
4	Takes little pleasure in activities	2.26	2.91
5	Lacks close friends	1.53	1.90
6	Seems indifferent to praise or criticism	1.35	4.02
7	Is emotionally cold with little affect	1.28	3.32

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



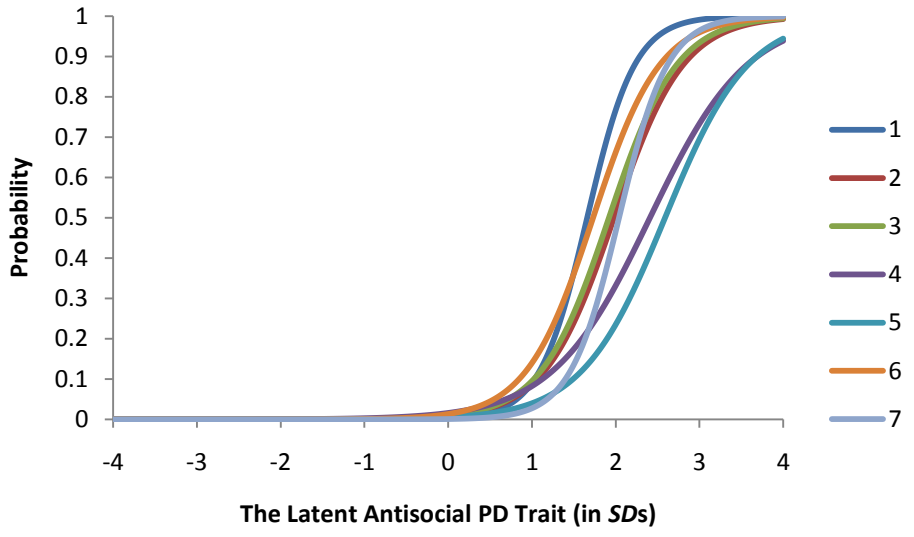
Antisocial Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Shows a failure to follow established laws and norms	3.50	1.66
2	Is deceitful	2.40	1.98
3	Is impulsive	2.45	1.92
4	Shows irritability and aggressiveness	1.71	2.41
5	Has a reckless disregard for safety	2.00	2.59
6	Shows consistent irresponsibility	2.49	1.73
7	Has a lack of remorse	3.43	2.04

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



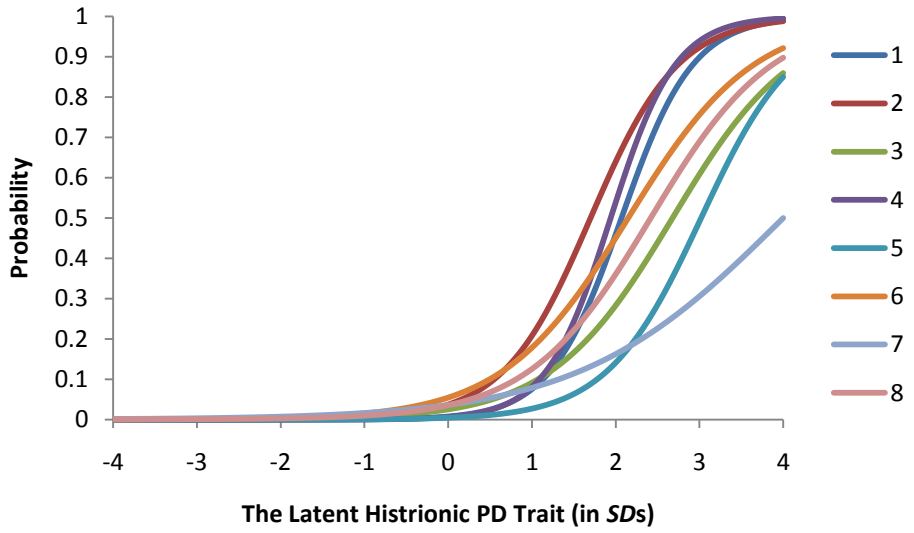
Histrionic Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Is uncomfortable when not the center of attention	2.34	2.07
2	Often interacts with others in a seductive way	1.91	1.70
3	Shallow and mercurial expression of emotion	1.37	2.68
4	Uses physical appears to draw attention	2.58	1.94
5	Uses impressionistic and vague speech	1.77	3.02
6	Has an exaggerated expression of emotion	1.33	2.15
7	Is suggestible	0.82	4.00
8	Considers relationships to be closer than actual	1.37	2.42

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



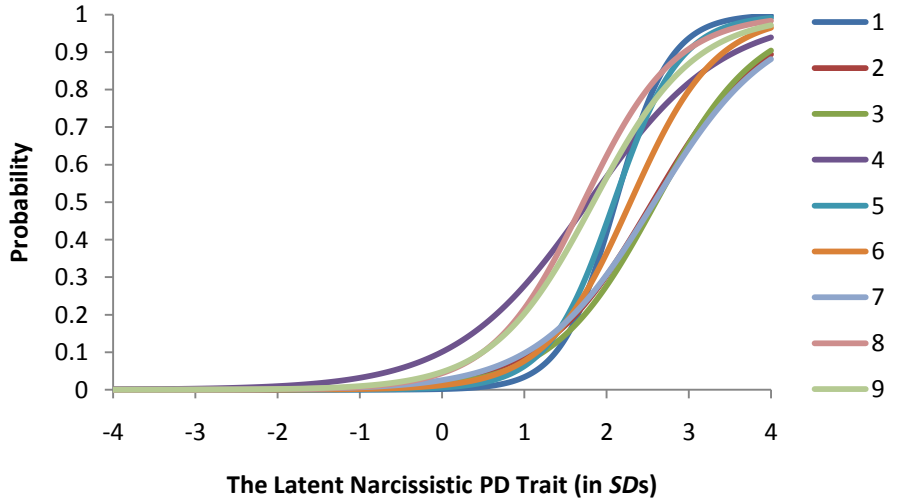
Narcissistic Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Has a grandiose sense of self	3.02	2.11
2	Has fantasies of power, beauty, etc.	1.48	2.56
3	Believes self is extraordinarily special	1.60	2.60
4	Requires admiration	1.23	1.78
5	Has a sense of entitlement	2.48	2.09
6	Is interpersonally exploitative	1.93	2.29
7	Lacks empathy	1.41	2.58
8	Is envious of others	1.79	1.72
9	Is arrogant or haughty	1.63	1.84

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



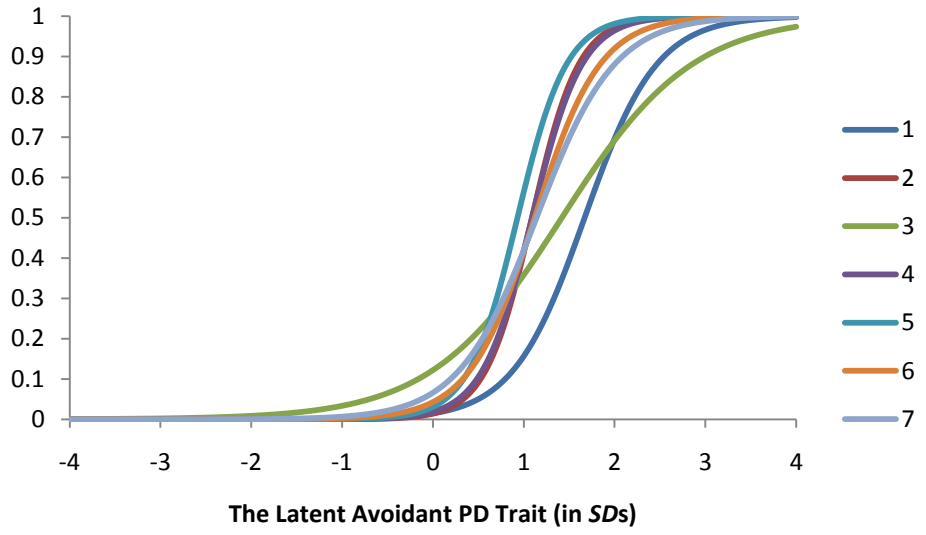
Avoidant Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Avoids occupational activities	2.51	1.67
2	Is afraid to get involved with others	3.88	1.09
3	Shows restraint for fear of being shamed	1.39	1.42
4	Is preoccupied with fears of criticism	3.62	1.09
5	Has fears of inadequacy in relationships	3.70	0.93
6	Views self as inept or unappealing	2.77	1.12
7	Is reluctant to take risks for fear of embarrassment	2.32	1.14

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



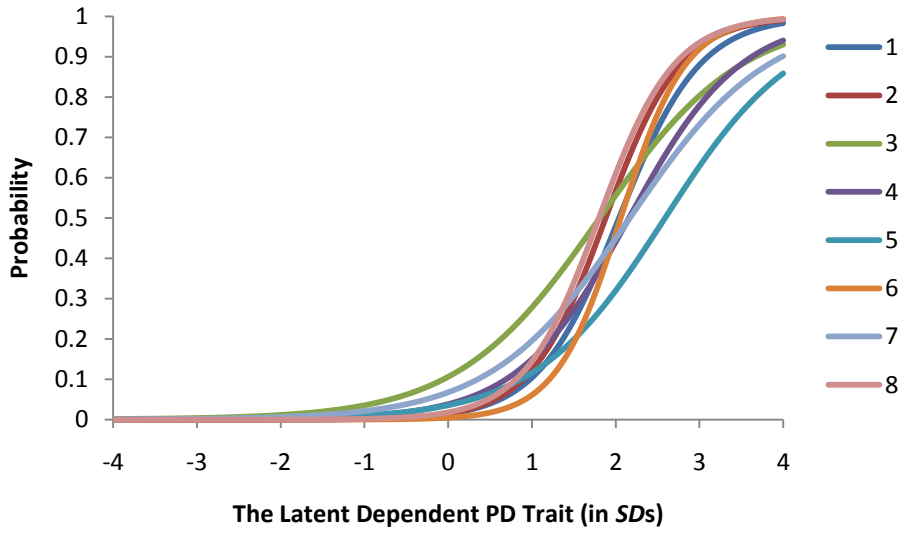
Dependent Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Has difficulty making decisions on one's own	2.07	2.04
2	Relies on others for major areas of life	2.23	1.88
3	Has trouble expressing disagreement	1.18	1.81
4	Has trouble doing things on one's own	1.50	2.16
5	Does excessive things for nurturance and support	1.28	2.59
6	Feels helpless when on one's own	2.58	2.06
7	Urgently seeks others' support	1.21	2.17
8	Is excessively afraid of having to take care of self	2.22	1.80

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



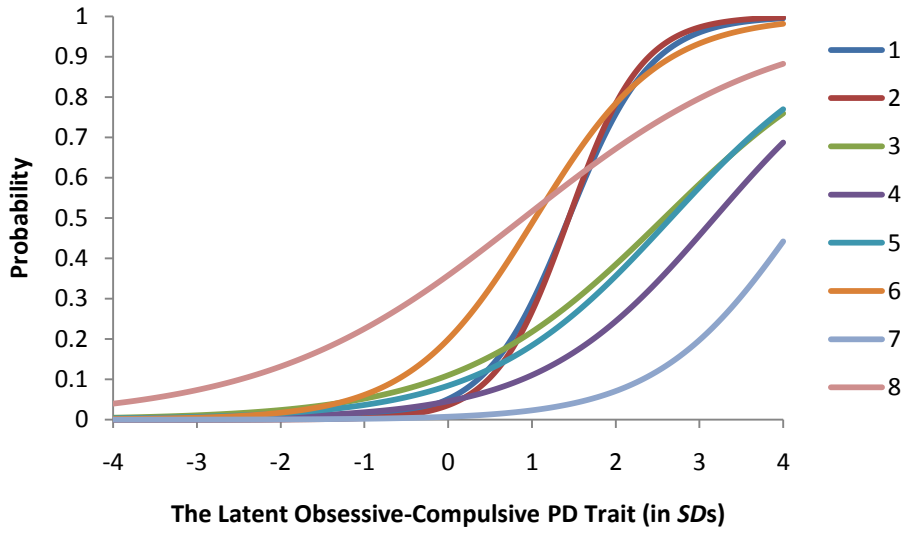
Obsessive-Compulsive Personality Disorder

Diagnostic Criteria and Item Parameters

#	Content	*a*	*b*
1	Is preoccupied with rules, details, and organization	2.03	1.44
2	Shows maladaptive perfectionism	2.29	1.44
3	Is excessively devoted to one's occupation	0.81	2.58
4	Is over-conscientious and inflexible	0.96	3.18
5	Is unable to discard worthless objects	0.90	2.66
6	Wants everything done in exactly one's own way	1.34	1.04
7	Has a miserly spending style	1.17	4.20
8	Is rigid and stubborn	0.65	0.90

Note. *a* = discrimination; *b* = difficulty.

Item Characteristic Curves



VITA

Name:	Luke D. Cooper
Address:	Department of Psychology Texas A&M University 4235 TAMU College Station, TX 77843-4235
Email Address:	lukedcooper@gmail.com
Education:	B.A., Psychology, Rowan University, 2007 M.S., Psychology, Texas A&M University, 2010