Capturing Abnormal Personality With Normal Personality Inventories: An Item Response Theory Approach

Kate E. Walton,[1] Brent W. Roberts,[2] Robert F. Krueger,[3] Daniel M. Blonigen,[3] and Brian M. Hicks[3]

[1]St. John's University

[2]University of Illinois, Urbana-Champaign

[3]University of Minnesota

ABSTRACT Correlational and factor-analytic methods indicate that abnormal and normal personality constructs may be tapping the same underlying latent trait. However, they do not systematically demonstrate that measures of abnormal personality capture more extreme ranges of the latent trait than measures of normal range personality. Item Response Theory (IRT) methods, in contrast, do provide this information. In the present study, we use IRT methods to evaluate the range of the latent trait assessed with a normal personality measure and a measure of psychopathy as one example of an abnormal personality construct. Contrary to the expectation that the measure of psychopathy would be more extreme than the measure of normal personality traits, the measures overlapped substantially in terms of the regions of the latent trait for which they provide information. Moreover, both types of inventories were limited in terms of measurement bandwidth, such that they did not provide information across the entire latent trait continuum. Implications and future directions are discussed.

The American Psychiatric Association's (1994) adherence to a medical model treats personality disorders as categorically distinct from

This article is based in part on a dissertation submitted by Kate E. Walton to the Department of Psychology, University of Illinois at Urbana-Champaign, in partial fulfillment of the requirements for the doctoral degree.

Correspondence should be addressed to Kate E. Walton, Department of Psychology, St. John's University, SB36 Marillac Hall, 8000 Utopia Parkway, Jamaica, NY 11439; E-mail: waltonk@stjohns.edu.

Journal of Personality 76:6, December 2008

DOI: 10.1111/j.1467-6494.2008.00533.x

one another as well as normal personality. However, personality disorders and abnormal personality constructs, such as the "Dark Triad" (Paulhus & Williams, 2002), are often thought of as extreme variants of normal personality traits. A large body of literature describes the relationship between normal personality models, such as the Five-Factor Model (e.g., Costa & Widiger, 2002), and abnormal personality. In general, these studies utilize a factor-analytic or correlational approach to appreciate the continuity between normal and abnormal personality. Given that abnormal personality constructs and personality disorders correlate with normal personality traits in predictable and consistent ways (e.g., Saulsman & Page, 2004), many have regarded such studies as evidence that abnormal personality constructs are simply exaggerations of normal traits. Our aim is to demonstrate how Item Response Theory methods can extend conclusions from factor-analytic and correlational approaches in fundamental ways.

Eysenck (1994) specifically demonstrated how correlational and factor-analytic techniques could be used to test what he called the "continuity hypothesis." As an example, assume one is interested in testing whether narcissism is continuous with normal personality and has a number of tests designed to measure narcissism. After administering the tests to narcissists and nonnarcissists, several analyses would be carried out to test the continuity hypothesis. First, the narcissism/nonnarcissism dichotomy would be correlated with the test scores to yield discrimination values where greater values indicate a greater ability of the test to differentiate between narcissists and nonnarcissists. Second, the intercorrelation matrix of the test scores would be factor-analyzed separately for narcissists and nonnarcissists, the resulting factor loadings serving as approximations of the correlations between the tests and the latent factor of narcissism. In order for the continuity hypothesis to be confirmed, two criteria must be met: The correlation between the factor loadings for the narcissistic and nonnarcissistic groups must approach unity, and the correlation between the discrimination values and the factor loadings (for both groups) must approach unity. Eysenck illustrated this with psychoticism and argued that these two criteria "would apply *if and only if* normality and psychosis form the extremes of a common continuum; if psychosis were a unique and separate disease, neither proportionality should apply" (1994, p. 9; italics in original).

While Eysenck's technique is indeed useful for demonstrating whether various forms of psychopathology lie on a common continuum with normal personality, factor-analytic methods offer only one piece of the puzzle. One significant piece that is missing indicates *where* on this common continuum these constructs lie, as we measure them, in relation to one another. While our theories and definitions of abnormal personality constructs place them at the far ends of continuous personality dimensions, factor-analytic and correlational methods do not formally test whether this definition is reflected in our measures. That is, given that we conceptualize personality disorders and subclinical forms of abnormal personality as constructs lying at the extreme ends of a continuum shared with normal personality traits, measures of psychopathology should contain items tapping more extreme ranges than measures of normal personality. This very issue was discussed more than a decade ago in an exchange concerning the use of normal personality inventories in clinical assessment. Costa and McCrae (1992) posed the question, "Are clinical and normal instruments redundant or complementary?" (p. 21), and Ben-Porath and Waller (1992) pointed out that "the ability of normal personality inventories to provide the breadth and magnitude of clinically relevant information that is found in clinical instruments ... remains to be demonstrated" (p. 23). To our knowledge, this has yet to be formally tested.

The primary objective of the current study is to demonstrate how an Item Response Theory (IRT) approach can be utilized to evaluate this. IRT models provide information concerning the region of the latent trait assessed by various types of inventories, those designed to assess normal range traits and those designed to assess abnormal constructs. In this study, we consider a measure of normal personality as well as a measure of psychopathy to illustrate this point. We first review some basic concepts associated with IRT (more technical discussions can be found elsewhere; e.g., Lord, 1980) and highlight how this approach can be advantageous in furthering our understanding of the nature of the relationship between normal and abnormal personality, insofar as how they are currently assessed.

An Overview of Item Response Theory

IRT includes various models that relate item characteristics with latent traits, represented by θ (theta), to predict test performance (Ham-

bleton, Swaminathan, & Rogers, 1991; Lord, 1980). While a number of different models can represent this relationship, we focus here on Samejima's Graded Response Model (GRM; Samejima, 1969), which is appropriate when item response options can be characterized as ordered categorical responses, such as Likert rating scales.

The relationship between theta and a person's response to an item can be described with option response functions (ORF). The GRM incorporates two parameters, *a* and *b*, that require explanation. The *a* parameter refers to item discrimination, the ability to differentiate among individuals at varying levels of theta. It is relative to the slope of the ORF; steeper slopes indicate greater discrimination power (see Embretson & Reise, 2000, for a more detailed explanation of the two-stage estimation procedure for obtaining the ORFs). The *b* parameters, often referred to as difficulty parameters, indicate the point along the theta scale at which respondents have a .50 probability of responding "above" that particular response category. Because we estimate the probability of responding above a particular response category, the number of difficulty parameters is one fewer than the number of response options ($j - 1$, where j = the number of response options). For example, suppose there are four response options and three (i.e., $j - 1$) *b* parameters. Assume the four response options correspond to *strongly disagree*, *disagree*, *agree*, and *strongly agree*. One *b* parameter would represent the point along the theta scale at which respondents have a .50 probability of responding above *strongly disagree* (i.e., responding with *disagree*, *agree*, or *strongly agree*). Panel A in Figure 1 depicts an example of ORFs for an item with four response options. In this example, the slope parameter, which is constant for each response category, is estimated to equal 1.58. This parameter determines the shape of the curves, such that higher *a* parameters will correspond to narrower, peaked curves, indicating good differentiation among individuals of different levels of theta. While the slope parameter determines the shape of the curves, the *b* parameters determine their location. There are three difficulty parameters estimated to be $-.43$, 1.14, and 2.47. Because the difficulty parameters are spread out across the upper end of latent trait continuum, the response functions are similarly spread out. Notice in Figure 1 that the curves peak in the middle of two adjacent *b* parameters (e.g., one curve peaks between 1.14 and 2.47; for more detailed information, consult Ostini & Nering, 2006).

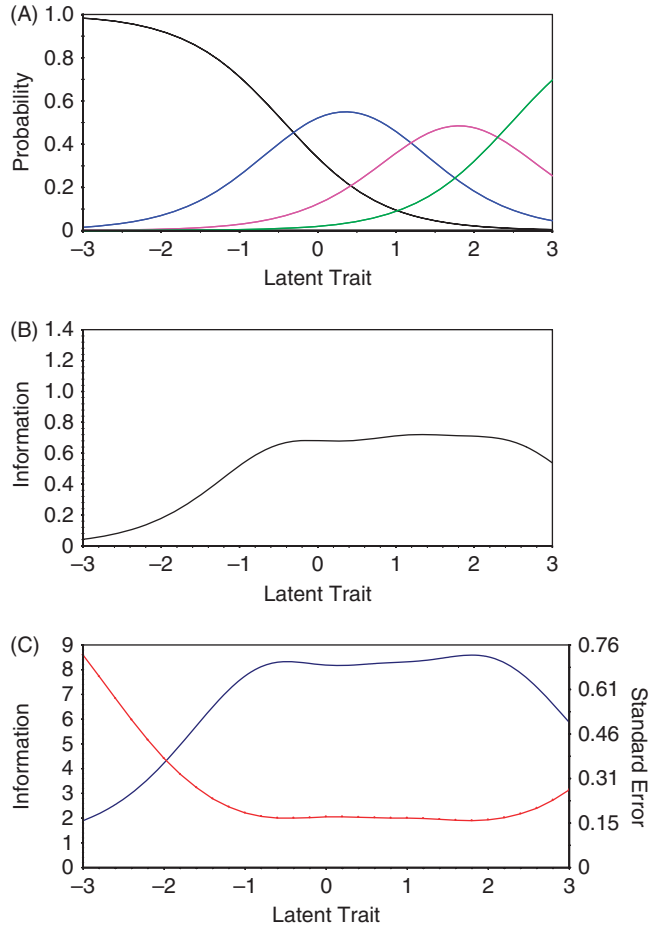

Figure 1
Example GRM option-response functions (Panel A), item-information function (Panel B), and test-information function (Panel C).

The ORFs can be transformed into item-information functions (IIFs), which indicate the amount of psychometric information an item contains across the entire span of the latent trait continuum (see Figure 1, Panel B). The amount of information an item provides is maximized around the *b* parameters. That is, IIFs peak at the level of the latent trait corresponding to the *b* parameters. In this particular example, the *b* parameters are spread out across the upper region of the latent trait, indicating that this item provides more precise measurement for individuals with high levels of the latent trait than for

those with low levels. If the *b* parameters were spread out across the entire continuum, the IIF would indicate that the item is fairly equal in precision for individuals with varying levels of the latent trait. An IIF is further influenced by the *a* parameter, such that items with better discriminating power provide more psychometric information. The curve in Panel B of Figure 1 is quite high due to the rather high discrimination parameter. The area under the curve would decrease in the event of a lower *a* parameter. One advantage of the IIFs is that they are additive and can be combined to form test information functions (TIF), which indicate the amount of information provided by an entire test. That is, we can determine how precise a test is at various levels of the latent trait. Notice in Panel C of Figure 1 that this particular test provides a significant amount of information across the entire spectrum but slightly less in the extreme regions. That is, it appears to consist of a number of items with fairly high *a* parameters and few extreme *b* parameters, particularly very low *b* parameters.

One other notable aspect of test information functions is that they have an inverse relationship with tests' standard errors of measurement, which, in an IRT framework, vary across the levels of the latent trait. If a test provides a significant amount of information at a particular level of theta, the standard error of measurement for individuals with levels of the latent trait within that region will be low. If an individual's level of theta is at a point where a test provides little or no information, his or her standard error of measurement would be high. The correspondence between the standard error of measurement and the amount of test information can be observed in Panel C of Figure 1 where the standard error is represented with a dotted line (see Embretson & Reise, 2000, for more details).

Item Response Theory in the Present Study

Although there are many advantages of using IRT over the traditionally used Classical Test Theory (see, e.g., Hambleton & Swaminathan, 1985), it is by virtue of the difficulty parameters and their effect on item- and test-information functions that IRT is particularly advantageous for our current purposes. As discussed above, correlational and factor-analytic studies provide some evidence that abnormal personality and certain dimensions in normal personality models may be tapping a common underlying latent trait. In an IRT framework, this is captured with the discrimination parameter, which

is analogous to a factor loading. Correlational designs, however, do not provide any information concerning the range of the underlying trait being assessed. IRT models that comprise difficulty parameters do, in contrast, provide this type of information. IRT models can systematically demonstrate whether measures of abnormal personality constructs, such as psychopathy, capture more extreme ranges of the latent trait than measures of normal range personality. As described above, this can most easily be detected by examining the TIFs, specifically, their location along the latent trait continuum. With this information, we can compare the psychometric information provided by abnormal and normal personality inventories. That is, we can determine whether the two types of measures are similar in terms of the amount of psychometric information (i.e., whether they have equally high discrimination parameters) and at which regions the inventories are most precise (i.e., the location of the difficulty parameter estimates). Moreover, the TIFs will enable us to determine whether both types of inventories offer precise measurement across the entire latent trait continuum, or whether there are regions of latent trait continuum that are assessed with less precision.

An Item Response Theory Investigation of Normal and Abnormal Personality Measures: The Current Study

In the present investigation, we employed IRT methods to investigate the psychometric properties of abnormal personality inventories relative to normal personality inventories. We examined a self-report psychopathy measure (the PPI; Lilienfeld & Andrews, 1996) as one example of abnormal personality, as well as a normal personality inventory (the MPQ; Tellegen, in press). Psychopathy was a likely candidate for this investigation given recent research aimed at understanding the relationship between psychopathy and normal personality traits. For example, Widiger and Lynam (1998) translated each characteristic of psychopathy (as defined by Hare's PCL-R; Hare, 1991, 2003) into the language of the Five-Factor Model, and Miller and his colleagues (Miller, Lynam, Widiger, & Leukefeld, 2001) consulted a number of psychopathy experts to derive a Big Five profile of the prototypical psychopath. Furthermore, correlational studies have demonstrated that psychopathy and domains of normal personality correlate in a predictable and consistent manner (Lynam & Derefinko, 2006). In terms of measures to use for the

current study, recent research suggested the PPI and MPQ as likely candidates. Benning and his colleagues have documented some overlap between the PPI and MPQ. Benning, Patrick, Hicks, Blonigen, and Krueger (2003) established a two-factor structure of the PPI and concluded that much of the variance in the factors could be captured with the MPQ. They further examined the criterion validity of the MPQ-estimated PPI factor scores and found that they exhibited a pattern of relationships with external variables that paralleled those previously reported for putatively direct measures of psychopathy (Benning, Patrick, Blonigen, Hicks, & Iacono, 2005).

With the PPI and MPQ being logical choices for investigation, in the current study we examined the similarity of their test information functions to determine, among other things, whether psychopathy inventories contain items tapping more extreme regions of the latent trait spectrum, as one might expect given that psychopathy is conceptualized as an extreme variant of normal personality dimensions.

METHOD

Participants

Participants were from the Minnesota Twin Registry, a birth-record-based registry of twins born in Minnesota. The sample used in the current study consists of male pairs born between the years of 1961 and 1964. Their ages ranged from 32 to 36 years ($M = 33$) at the time of assessment. The sample consists of 89 monozygotic pairs, 47 dizygotic pairs, and 81 individuals whose co-twin did not participate (total $N = 353$). Additional details concerning the Minnesota Twin Registry and the particular sample used in the following analyses have been described elsewhere (Blonigen, Carlson, Krueger, & Patrick, 2003; Krueger & Johnson, 2002; Lykken, Bouchard, McGue, & Tellegen, 1990).

Measures

The Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996) was used to assess psychopathy. The PPI is a self-report measure designed to assess the core personality characteristics associated with psychopathy. Its subscales include Machiavellian egocentricity, social potency, fearlessness, cold-heartedness, impulsive nonconformity, blame externalization, carefree nonplanfulness, and stress immunity. In total, the PPI consists of 187 items with four response options, including *false*,

mostly false, *mostly true*, and *true*. The eight subscales all had high internal consistency, with alphas ranging from .75 to .89 ($M = .81$).

The 198-item version of the Multidimensional Personality Questionnaire (MPQ; Tellegen, 1982, in press) was administered as a measure of normal personality. The MPQ is comprised of 11 subscales which tap three higher order dimensions, including positive emotionality, negative emotionality, and constraint. The 11 primary scales include well-being, social potency, achievement, social closeness, stress reaction, alienation, aggression, self-control, harm avoidance, traditionalism, and absorption. There were four response options for each item (*definitely true*, *probably true*, *probably false*, and *definitely false*). The 11 primary scales had high internal consistency, with alphas ranging from .82 to .92 ($M = .86$).

Procedure

We first determined which items from the MPQ were most highly related to the eight PPI scales. After confirming the unidimensionality of the scales, the items comprising the MPQ and PPI scales were then analyzed in an IRT framework to determine whether they provide comparable amounts of information across the latent trait. In the analyses detailed below, the PPI scales and the items from the MPQ selected to represent these scales were analyzed separately as well as jointly. Although our main objective was to compare the psychometric information across the two scales, combining the items from the two scales and analyzing them together was essential to ensure that the items from the two scales could be considered to be tapping the same latent trait and could therefore be meaningfully compared. Several steps were taken throughout the procedures to address this issue and are detailed below.

ANALYSES AND RESULTS

Item Selection

It was first necessary to determine which items could be drawn from the MPQ to represent the eight PPI scales. To this end, the total scale scores from the PPI were correlated with each item from the MPQ. We selected the MPQ items with the highest correlations with the PPI scales, though the number of MPQ items selected could not exceed the number of items within each PPI scale. The amount of information a scale provides is directly influenced by the number of items comprising it (provided the items are good markers of the latent trait); therefore, we kept the scales equal in this respect. Two

additional stipulations were required for item selection: First, the items selected were those exhibiting the highest correlation with their target PPI scale (e.g., each of the MPQ items selected to represent fearlessness correlated more highly with the fearlessness scale score than with any of the other seven PPI scale scores). Second, the correlations had to reach at least .25. Meeting these criteria were MPQ items to represent the PPI scales of stress immunity, blame externalization, social potency, and carefree nonplanfulness.

For the remaining four PPI scales, there were a smaller number of MPQ items, eight or fewer, that met the criteria. In one of these cases, the fearlessness scale, we selected eight items from the PPI scale to compare with the eight MPQ items selected to represent fearlessness. We chose the eight items with the highest loadings in a single-factor solution. We employed this strategy for the seven MPQ items that met the criteria for representing Machiavellian egocentricity as well. However, this method was not satisfactory in this case because combining the PPI and MPQ items did not produce a sufficiently unidimensional structure. The PPI and MPQ items loaded highly on two separate factors. Because Machiavellian egocentricity is central to psychopathy and we believed a core component of it could be captured by the MPQ aggression scale, we employed a different strategy for selecting items. The PPI Machiavellian egocentricity and the MPQ aggression items were entered into a principal components analysis together. The scree plot suggested two or four components. Two-, three-, and four-component models were then considered. The three- and four-component solutions did not appear to provide a good fit, as the third and fourth components were comprised of very few items, and there were many high cross-loadings. Therefore the two-component solution was chosen. Based on that solution, we selected PPI and MPQ items that loaded .35 or greater on the primary component. This yielded 16 MPQ items (all but two) and 14 PPI items. Two of these MPQ items (those with the lowest loadings) were not included in order to keep the number of items equal.

Additional steps (e.g., a principle component analysis with PPI and MPQ items) were not taken to recover a set of MPQ items to represent cold-heartedness and impulsive nonconformity as very few MPQ items correlated highly with the scales. In the case of impulsive nonconformity, only four items reached a correlation of .25 the PPI scale. Only seven MPQ items met both criteria for cold-heartedness (although there were a few additional items that reached a correla-

tion of .25 with cold-heartedness, they had higher correlations with another scale). Furthermore, the cold-heartedness scale has been found to be problematic in prior research (Benning et al., 2003). The difficulty we encountered in recovering MPQ items to represent these two scales is discussed in greater detail later.

The internal consistency of the MPQ-estimated blame externalization, carefree nonplanfulness, social potency, stress immunity, fearlessness, and Machiavellian egocentricity scales reached .92, .84, .90, .85, .77, and .85 respectively. The internal consistency of the eight fearlessness items selected from the PPI reached .80, and the internal consistency of the 14 PPI Machiavellian egocentricity items reached .76. The MPQ items were combined with their respective PPI items (e.g., the items from the PPI blame externalization scale and the items from the MPQ selected to represent blame externalization were combined), and the reliability of these joint scales was assessed as well. Alphas reached .94, .88, .94, .88, .86, and .87. The remaining two scales (cold-heartedness and impulsive nonconformity) were not considered in the subsequent analyses as we were unable to identify MPQ items to represent them.

Dimensionality

The GRM is most easily interpreted when the items of a scale tap a single dimension. To determine if the items were unidimensional, we conducted an exploratory factor analysis (EFA) of polychoric correlations for each scale using Mplus (Muthén & Muthén, 2004) and considered the scales sufficiently unidimensional if the root mean square residual (RMSR) did not exceed .10 (Kline, 2005). In all but one case, the RMSR did not exceed .10. The RMSR for the MPQ-estimated carefree nonplanfulness scale failed to meet the criterion set. Two items from the scale with a strong loading on a second factor were identified and were eliminated from a second EFA. The RMSR improved and did not exceed .10. Next, we combined each PPI scale with its MPQ complement and carried out an EFA to determine whether the scales from the PPI and MPQ could be considered to be tapping the same latent trait. In each case, the RMSR did not exceed .10. Finally, to confirm that a one-factor model fit the MPQ, PPI, and combined scales reasonably well, we carried out a confirmatory factor analysis for each scale using Mplus and examined several fit indices including the Tucker Lewis Index, the root

Table 1
Items Used in IRT Analyses

Scale	Item Numbers
Blame Externalization	
PPI	23, 30, 33, 36, 67, 80, 82, 92, 99, 105, 113, 138, 141, 161, 162, 165, 167, 176
MPQ (300-item version)	12, 29, 57, 74, 88, 103, 133, 147,161, 178, 192, 205, 226, 238, 250, 268, 283, 298
MPQ (198-item version)	9, 17, 34, 42, 51, 59, 80, 88, 101, 114, 122, 131, 148, 158, 166, 176, 189, 198
Carefree Nonplanfulness	
PPI	4, 10, 15, 48, 56, 62, 64, 68, 77, 87, 97, 112, 120, 148, 164, 177, 183, 184
MPQ (300-item version)	2, 10, 26, 41, 51, 90, 98, 104, 110, 162, 164, 176, 190, 196, 214, 228, 239, 282
MPQ (198-item version)	2, 8, 15, 27, 31, 53, 56, 60, 65, 102, 103, 112, 121, 125, 138, 149, 159, 188
Machiavellian Egocentricism	
PPI	20, 38, 39, 44, 75, 110, 122, 133, 137, 143, 150, 170, 173, 179
MPQ (300-item version)	7, 22, 66, 82, 97, 112, 127, 158, 172, 202, 217, 232, 261, 293
MPQ (198-item version)	5, 13, 38, 47, 55, 67, 75, 98, 109, 128, 141, 152, 173, 196
Social Potency	
PPI	1, 3, 7, 9, 14, 18, 22, 31, 35, 41, 49, 54, 55, 69, 72, 90, 102, 114, 131, 139, 149, 155, 157, 185
MPQ (300-item version)	1, 25, 34, 47, 56, 60, 71, 86, 94, 105, 113, 117, 150, 187, 206, 209, 218, 221, 233, 244, 256, 257, 278, 287
MPQ (198-item version)	1, 14, 22, 30, 33, 35, 40, 50, 54, 61, 68, 70, 91, 119, 132, 134, 142, 143, 153, 162, 169, 170, 184, 191
Fearlessness	
PPI	5, 26, 34, 42, 59, 142, 154, 181
MPQ (300-item version)	77, 124, 128, 149, 171, 204, 225, 237
MPQ (198-item version)	45, 74, 76, 90, 108, 130, 147, 157
Stress Immunity	
PPI	6, 60, 63, 73, 86, 117, 121, 136, 144, 160, 169

(Continued)

Table 1 (Cont.)

Scale	Item Numbers
MPQ (300-item version)	3, 15, 27, 72, 107, 131, 154, 175, 188, 222, 270
MPQ (198-item version)	3, 10, 16, 41, 63, 79, 95, 111, 120, 144, 177

mean square error of approximation, and the standardized root mean square residual. All indices indicated that the scales were satisfactorily unidimensional. The final set of items used in the IRT analyses detailed later can be found in Table 1, which includes the item numbers as they appear in the PPI and the MPQ (the full 300-item version, as well as the 198-item version).

Item Parameter Estimations

Fitting a GRM, the items' discrimination and difficulty parameters were estimated with MULTILOG (Thissen, 2003), and the test information functions were plotted. In order to ensure that the PPI scales and their MPQ counterparts were on the same metric in terms of the distribution of theta (the mean and standard deviation of which is normally set to zero and unity, respectively, for identification purposes), the parameters for the PPI and MPQ items were estimated jointly. Then the TIFs were calculated separately for the PPI and MPQ scales using the jointly calibrated item parameter estimates.

A visual interpretation of TIFs is often used as a means to evaluate the similarity of two or more scales, and scales are considered parallel if they have nearly identical TIFs (Stark, Chernyshenko, Chuah, Lee, & Wadlington, 2001). The PPI and MPQ TIFs are presented in Figure 2. By and large, the TIFs appear to be very similar across the two measures. The PPI blame externalization scale and the corresponding MPQ items both provide the most amount of information at the high end of the latent trait while their TIFs decline below $-.5$. A similar pattern is observed for the PPI carefree nonplanfulness scale and its MPQ counterpart. Likewise, the PPI and MPQ Machiavellian egocentricity scales provide a lot of information in the upper region of the latent trait, and their TIFs decline at approximately $+.5$. The blame externalization, carefree non-

planfulness, and Machiavellian egocentricity scales appear to have few items tapping the moderately low and extremely low regions of the latent trait. Instead, there are a plethora of items with moderately and extremely high difficulty parameters. The PPI social potency scale and the MPQ items that capture social potency are fairly consistent in terms of the amount of information they provide across the latent

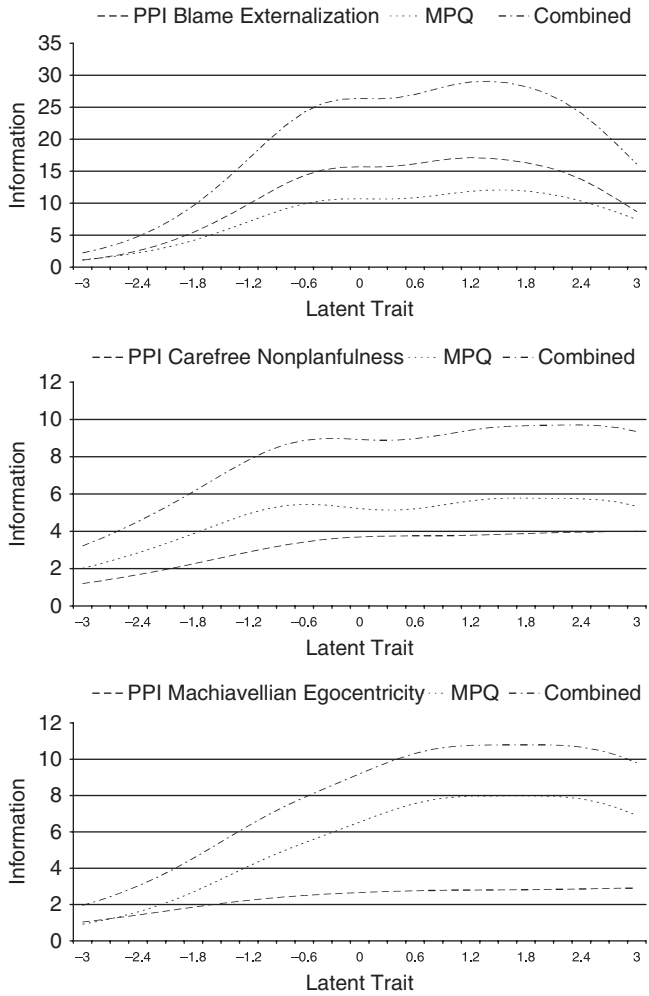

Figure 2
GRM test-information functions for the PPI scales, the MPQ-estimated PPI scales, and the combined PPI and MPQ scales.

trait. They provide a lot of information in the moderate range of the latent trait, and their TIFs taper off a bit at both extremes. Less information is provided for individuals with latent trait estimates exceeding ± 2.0. The pattern is more extreme for the fearlessness scales, which are bell-shaped, peaking in the middle of the latent trait and declining fairly steeply at both extremes. In contrast, the stress immunity scales both consist of a significant amount of items tapping the low and moderate regions but contain few items with high difficulty parameters. The TIF declines at approximately +1.0.

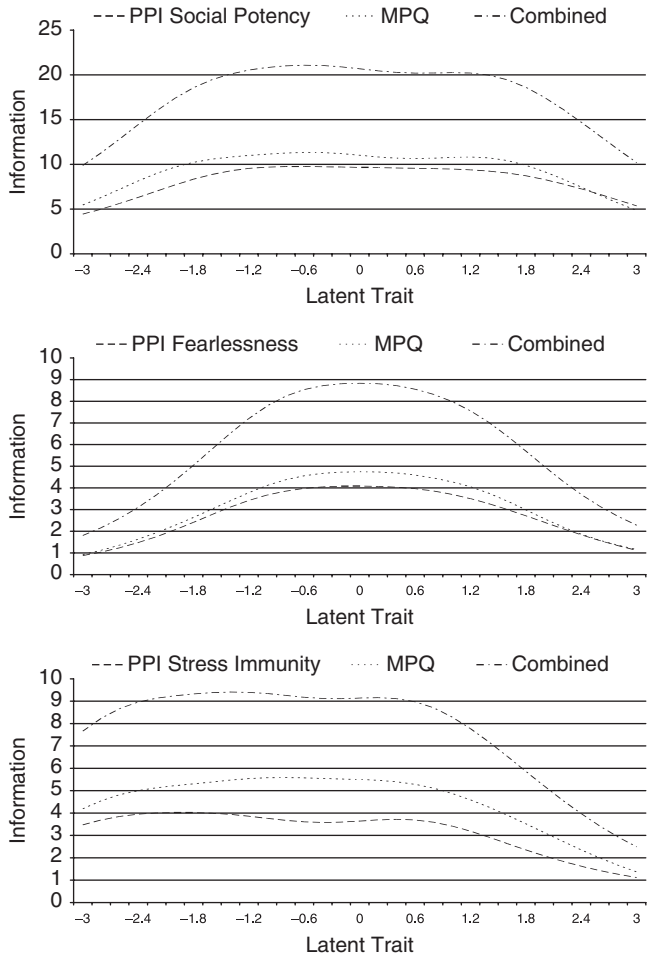

Figure 2
Continued.

To supplement the graphical representation of the similarity in TIFs across the two measures, we calculated the ratio of their information functions, a method used previously to estimate the efficiency of two scales at various points along the latent trait continuum (Cooke & Michie, 1997; Cooke, Michie, Hart, & Hare, 1999; Nunnally & Bernstein, 1994). The information values for the two measures, as well as the ratio of the scales' information values at various points along the latent trait continuum can be found in Table 2. Remarkably, for all but one scale, the MPQ items provided more information than the PPI items across the entire continuum. The blame externalization domain was the only instance in which the PPI items provided more information.

The TIFs for the joint scales also appear in Figure 2. As expected, they are highly similar in shape to those obtained when inspecting the PPI and MPQ items separately. Moreover, the location of the joint TIF did not differ from that of the separate TIFs. The difference between the separate scales' TIFs and the TIFs for the combined scales lies only in the amount of information provided across the spectrum, such that combining the two scales yields more information, and thus more accurate estimates of individuals' latent traits, across the spectrum.

DISCUSSION

The objective of the current study was to demonstrate how Item Response Theory methods can be used to evaluate systematically whether measures of abnormal personality constructs capture more extreme ranges of the latent trait than measures of normal range personality or whether measures of normal and abnormal personality assess distinct regions of a common continuum. In this study we selected one abnormal personality construct on which to focus, namely psychopathy, for illustrative purposes. Previous researchers have delineated the relationship between abnormal personality constructs, such as psychopathy, and normal personality traits in factor-analytic and correlational studies. While these methods suggest that abnormal personality constructs may indeed be related to a particular constellation of normal personality dimensions, they do not provide any information concerning the level of the latent trait the respective

Table 2
Information for PPI and MPQ-Estimated PPI Scales, and the Ratio of the Two at Various Levels of Theta

	Theta										
Scale	−3.0	−2.4	−1.8	−1.2	−.6	0	+.6	+1.2	+1.8	+2.4	+3.0
Blame Externalization (Blame)	1.08	2.54	5.43	9.85	14.33	15.67	16.15	17.10	16.33	13.72	8.68
MPQ-estimated Blame	1.14	2.22	4.17	7.11	9.90	10.69	10.84	11.85	11.90	10.32	7.44
MPQ/PPI	1.06	0.87	0.77	0.72	0.69	0.68	0.67	0.69	0.73	0.75	0.86
Carefree Nonplanfulness (Care)	1.20	1.68	2.26	2.89	3.43	3.70	3.76	3.79	3.88	3.96	4.00
MPQ-estimated Care	2.01	2.84	3.87	4.94	5.45	5.22	5.20	5.63	5.78	5.74	5.34
MPQ/PPI	1.68	1.69	1.71	1.71	1.59	1.41	1.38	1.49	1.49	1.45	1.34
Machiavellian Egocentricism (Mach)	1.04	1.43	1.86	2.24	2.50	2.66	2.76	2.80	2.82	2.86	2.90
MPQ-estimated Mach	0.91	1.59	2.69	4.11	5.39	6.54	7.56	7.96	7.98	7.81	6.91
MPQ/PPI	0.88	1.11	1.45	1.83	2.16	2.46	2.74	2.84	2.83	2.73	2.38
Social Potency (SP)	4.43	6.30	8.34	9.55	9.75	9.67	9.56	9.39	8.72	7.25	5.35
MPQ-estimated SP	5.45	8.13	10.23	10.99	11.33	11.01	10.64	10.82	9.89	7.45	4.83
MPQ/PPI	1.23	1.29	1.23	1.15	1.16	1.14	1.11	1.15	1.13	1.03	0.90
Fearlessness (Fear)	0.88	1.47	2.41	3.42	3.98	4.09	3.97	3.51	2.71	1.83	1.12
MPQ-estimated Fear	0.92	1.63	2.64	3.78	4.57	4.75	4.59	4.05	2.99	1.86	1.16
MPQ/PPI	1.05	1.11	1.10	1.11	1.15	1.16	1.16	1.15	1.10	1.02	1.04
Stress Immunity (Stress)	3.47	3.95	4.02	3.86	3.61	3.64	3.69	3.19	2.35	1.62	1.12
MPQ-estimated Stress	4.19	5.01	5.31	5.54	5.57	5.50	5.28	4.59	3.52	2.35	1.37
MPQ/PPI	1.21	1.27	1.32	1.44	1.54	1.51	1.43	1.44	1.50	1.45	1.22

inventories assess. In contrast, we used modern psychometrics, IRT, which have the capability of providing this insight. We fit an IRT model to the items from two scales, one normal personality inventory and one measure of psychopathy, to determine if they provide comparable amounts of psychometric information and the regions of the latent trait at which they provide the most information.

We were able to select items from the normal personality inventory, the MPQ, to represent six of the eight scales of the psychopathy measure, the PPI. We concluded that these six PPI scales and their MPQ counterparts measure the same regions of the latent trait with comparable precision. The test information functions for each measure were highly similar both in terms of how informative they were, as well as at which particular regions they were most informative. None of the scales was uniformly consistent across the latent trait continuum in terms of how informative they were, as some offered little information at both extremes (i.e., very low and very high levels of the latent trait), while others offered little information in only one extreme region. Furthermore, combining items from the two instruments proved to yield similarly shaped TIFs and more informative measures. This reflects the fact that the discrimination parameters were not altered much when the items from the two measures were combined, indicating that the PPI and MPQ scales are comprised of items tapping the same constructs. This also reflects the stability of the difficulty parameters, indicating that when placed on the same scale, the PPI and MPQ items still cover the same region of the latent trait. The increased amount of information obtained by combining the scales would enable better discriminations among individuals at varying levels of the latent trait, would increase scale reliability, and would ostensibly lead to better predictive validity. The primary conclusion based on these findings suggests that measures of normal range personality capture much of the information being obtained with a "direct" measure of psychopathy. Several of the PPI's scales have counterparts in the MPQ, and the PPI scales provide no psychometric information above and beyond that available with the MPQ.

It should not go unmentioned, though, that not all PPI scales could be represented with MPQ items. The scales tapping cold-heartedness and impulsive nonconformity could not be recreated with MPQ items. We can speculate as to why this occurred. The difficulties with the cold-heartedness scale have been discussed pre-

viously. Benning et al. (2003) reported a failure of the cold-heartedness scale to load on either of the two PPI factors (the first of which is characterized by low anxiousness, social dominance, and risk taking, and the second of which is characterized by impulsiveness, unconventionality, aggression, and estrangement from others). Benning et al. speculated that the scale may be limited in its ability to capture the affective deficits thought to be central components of psychopathy. Instead, the items reflect sentimentality and imaginativeness. In terms of the MPQ scales, cold-heartedness is most highly related to absorption, and to a lesser extent, stress reaction. Absorption, however, is not related to psychopathy. Benning et al. (2003) concluded that the PPI cold-heartedness scale captures characteristics not associated with the other PPI scales and perhaps unrelated to psychopathy. It should be noted, however, that Benning and colleagues' findings were based on the same sample utilized in the current study. Therefore, it is possible that the difficulties with the cold-heartedness scale are specific to this sample. Furthermore, given prior evidence that the cold-heartedness scale is related to domains of the Five-Factor Model, the problem could lie with the MPQ rather than the cold-heartedness scale itself. For example, cold-heartedness has been shown to be negatively related to the agreeableness facets of altruism and tender-mindedness and the extraversion facets of warmth and positive emotions (Derefinko & Lynam, 2006), which are central aspects of psychopathy, This might suggest that the MPQ is limited in its capacity to capture some key components of psychopathy. Finally, although one might expect the PPI impulsive nonconformity scale to correlate highly with items from the MPQ traditionalism scale, the item content across the two scales differs slightly. The PPI impulsive nonconformity scale includes items tapping impulsivity versus constraint (e.g., preferring to have vacations carefully planned out), as well as items tapping rebelliousness (e.g., open to the idea of wearing one's hair in a mohawk). The MPQ traditionalism scale focuses more on old-fashioned, traditional values (e.g., indicating that in raising one's children, high morals and values are of primary importance).

In short, in comparing the test information functions generated from our IRT parameter estimations, we determined that psychopathy and normal personality measures are quite redundant in terms of the information they provide. Furthermore, both measures do not provide information to an equal extent across the entire range of

the latent trait. In most cases, greater information was provided in the moderate regions of the trait with less information in the extreme regions. The implications for these findings will be discussed in turn.

Implications

First, we turn to the general issue concerning the relationship between normal and abnormal personality. We successfully fit an IRT model assuming unidimensional items and a continuous latent trait, and although in the present studies we considered only one type of personality disorder—psychopathy—we believe that our findings provide further and more definitive evidence that there is continuity between normal and abnormal personality and that "individual differences in personality disorder are best represented by a dimensional system" (Livesley, 2001, p. 278).

Of course, this has implications not only for how to conceptualize personality disorders but for the measurement of abnormal personality as well. In the present study we demonstrated that normal personality inventories can capture many of the interpersonal, affective, and behavioral aspects of one abnormal personality construct, psychopathy, and that no psychometric information was offered by the psychopathy inventories above and beyond that provided by the normal personality inventories. In fact, in all but one case, the MPQ items actually provided more information than their corresponding PPI items. We find this quite provocative. At the very least, this finding suggests that our inventories are not necessarily assessing what they are intended to measure. Normal personality inventories are intended to provide significant information in the moderate range of the trait continuum, while abnormal personality inventories are intended to provide information about their target respondent, consisting of items endorsed only by those extremely high on the latent trait (Cooke & Michie, 1997). This appears, however, not to be the case, and, to our knowledge, this is the first systematic evaluation of how well normal personality inventories can capture the breadth and magnitude of clinically relevant information (Ben-Porath & Waller, 1992). It may seem that our classification of inventories as "normal" or "abnormal" up to this point has been somewhat arbitrary. O'Connor (2002) was quite insightful when he wrote, "The categorization of inventories for their focus on normal personality or psychopathology is not always an obvious or straightforward

exercise'' (p. 974). In light of our findings, it is fair to conclude that our existing measures of normal and abnormal personality are redundant, as Costa and McCrae (1992) pondered.

Another finding to consider is that pertaining to the limited coverage of the inventories across the latent trait continuum. In a number of previous studies the implications for limited measurement bandwidth have been discussed. Haigler and Widiger (2001) discussed the failure to observe many relationships between normal personality traits and personality disorders that one might expect. For example, they pointed to the lack of significant correlations found between conscientiousness and obsessive-compulsive personality disorder. They hypothesized and tested the idea that the lack of such expected correlations is due to the fact that normal personality measures do not contain a great deal of extreme items, particularly in the adaptive direction (e.g., extremely high conscientiousness). Only after altering the item content of a normal personality measure to include more extreme items did Haigler and Widiger observe the correlations they expected to find. Our inability to identify MPQ items that are highly related to certain psychopathy scales could be explained by the narrow range of the latent trait captured by both the MPQ and PPI items.

Limited measurement bandwidth can also have implications for assessing development. For example, if a researcher is using a psychopathy measure that provides little information at the very high end of the trait, the ability to assess the development of a highly psychopathic person would be limited. Furthermore, Fraley, Brennan, and Waller (2000) empirically showed that limited measurement bandwidth can lead to inflated test-retest correlations, resulting in erroneous estimates of continuity. In addition to the failure to detect expected relationships among constructs and problematic estimates of development, limited measurement bandwidth can lead to problems in selection such that a researcher may not be able to distinguish among individuals at extremely high estimates of the latent trait.

Finally, Walton (2005) demonstrated that limited measurement bandwidth can lead to erroneous detection of a taxon in a taxometric analysis. She simulated a data set in which the hypothetical participants assumed a continuous distribution and the items' difficulty parameters did not span the entire continuum. Despite the continuous distribution of the participants, a low base rate taxon was detected.

Limitations and Future Directions

Although we find the results of this study compelling, we must highlight a few limitations and discuss possible future directions. First, we focused solely on psychopathy. Prior to concluding definitively that normal and abnormal personality lie on a common continuum and that measures of normal personality can be used to assess personality disorder, the methods we have employed here should be applied to other personality disorders as well. Furthermore, although we were successful in generating indices of several psychopathy scales with items from normal personality measures, not all aspects of psychopathy could be located within normal personality inventories. Perhaps this is a limitation of the particular measure of normal personality employed here, or perhaps there are aspects of personality disorder not located in any omnibus personality inventories. Alternatively, the psychopathy scales that could not be captured with normal personality inventories could be criticized for not measuring certain elements of psychopathy well.

It should also be noted that our findings may be limited to self-report measures of personality disorder. In the realm of psychopathy, for example, it would be of interest to see if our findings generalize to non-self-report measures, such as Hare's PCL-R. Based on the work of Cooke and colleagues (Cooke & Michie, 1997; Cooke et al., 1999), it appears as though the PCL-R lacks items tapping the extreme regions of the latent trait, though a systematic comparison between the PCL-R and normal personality inventories has yet to be carried out. Finally, future work should aim at revising existing inventories or creating new inventories that successfully capture the entire spectrum of the latent trait of interest. In most cases, this would entail generating items with extreme difficulty parameters.

CONCLUSIONS

This work was carried out to help inform the study of personality and psychopathology and to provide the means to test the relative efficiency of normal and abnormal personality inventories. We have demonstrated the utility of Item Response Theory in formally comparing the ability of normal and abnormal personality inventories to assess disordered personality. In addition to providing strong evidence for the continuity between normal and abnormal traits and

evidence for the great degree of psychometric overlap between normal and abnormal measures, the measures are somewhat limited in terms of measurement bandwidth. It is our hope that our findings will not only lead to the revision of inventories to assess the full range from moderate and normal to extreme and abnormal, but will also highlight the need to rely on modern psychometrics to inform not only the measurement of personality and psychopathology but the relationship between the two as well.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Ben-Porath, Y. S., & Waller, N. G. (1992). Five big issues in clinical personality assessment: A rejoinder to Costa and McCrae. *Psychological Assessment*, **4**, 23–25.

Benning, S. D., Patrick, C. J., Blonigen, D. M., Hicks, B. M., & Iacono, W. G. (2005). Estimating facets of psychopathy from normal personality traits. *Assessment*, **12**, 3–18.

Benning, S. D., Patrick, C. J., Hicks, B. M., Blonigen, D. M., & Krueger, R. F. (2003). Factor structure of the Psychopathic Personality Inventory: Validity and implications for clinical assessment. *Psychological Assessment*, **15**, 340–350.

Blonigen, D. M., Carlson, S. R., Krueger, R. F., & Patrick, C. J. (2003). A twin study of self-reported psychopathic personality traits. *Personality and Individual Differences*, **35**, 179–197.

Cooke, D. J., & Michie, C. (1997). An Item Response Theory of the Hare Psychopathy Checklist – Revised. *Psychological Assessment*, **9**, 3–14.

Cooke, D. J., Michie, C., Hart, S. D., & Hare, R. D. (1999). Evaluating the screening version of the Hare Psychopathy Checklist – Revised (PCL:SV): An Item Response Theory analysis. *Psychological Assessment*, **11**, 3–13.

Costa, P. T., Jr., & McCrae, R. R. (1992). Reply to Ben-Porath and Waller. *Psychological Assessment*, **4**, 20–22.

Costa, P. T., Jr., & Widiger, T. A. (2002). *Personality disorders and the Five-Factor Model of Personality*. Washington, DC: American Psychological Association.

Derefinko, K. J., & Lynam, D. R. (2006). Convergence and divergence among self-report psychopathy measures: A personality-based approach. *Journal of Personality Disorders*, **20**, 261–280.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Eysenck, H. J. (1994). Normality-abnormality and the Three-Factor Model of Personality. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality* (pp. 3–25). New York: Springer Publishing Company, Inc.

Fraley, R. C., Brennan, K. A., & Waller, N. G. (2000). An Item Response Theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, **78**, 350–365.

Haigler, E. D., & Widiger, T. A. (2001). Experimental manipulation of NEO-PI-R items. *Journal of Personality Assessment*, **77**, 339–358.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and application*. Boston: Kluwer-Nijhoff.

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

Hare, R. D. (1991). *The Hare Psychopathy Checklist – Revised*. Toronto: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist – Revised* (2nd ed.). Toronto: Multi-Health Systems.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Krueger, R. F., & Johnson, W. (2002). The Minnesota Twin Registry: Current status and future directions. *Twin Research*, **5**, 488–492.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, **66**, 488–524.

Livesley, W. J. (2001). Commentary on reconceptualizing personality disorder categories using trait dimensions. *Journal of Personality*, **69**, 277–286.

Lord, F. M. (1980). *Applications of Item Response Theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lykken, D. T., Bouchard, T. J., Jr., McGue, M., & Tellegen, A. (1990). The Minnesota twin family registry: Some initial findings. *Acta Geneticae Medicae et Gemellologicae*, **39**, 35–70.

Lynam, D. R., & Derefinko, K. J. (2006). Psychopathy and personality. In C. Patrick (Ed.), *Handbook of psychopathy* (pp. 133–155). New York: Guilford.

Miller, J. D., Lynam, D. R., Widiger, T. A., & Leukefeld, C. (2001). Personality disorders as extreme variants of common personality dimensions: Can the Five-Factor Model adequately represent psychopathy? *Journal of Personality*, **69**, 253–276.

Muthén, L. K., & Muthén, B. O. (2004). Mplus (Version 3.0) [Computer software]. Los Angeles: Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

O'Connor, B. P. (2002). The search for dimensional structure differences between normality and abnormality: A statistical review of published data on personality and psychopathology. *Journal of Personality and Social Psychology*, **83**, 962–982.

Ostini, R., & Nering, M. L. (2006). *Polytomous Item Response Theory models*. Thousand Oaks, CA: Sage Publications.

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, **36**, 556–563.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, Monograph Supplement, No. 17.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review*, **23**, 1055–1085.

Stark, S., Chernyshenko, S., Chuah, D., Lee, W., & Wadlington, P. (2001). *IRT modeling lab*. Retrieved June, 16, 2006, from http://work.psych.uiuc.edu/irt/

Tellegen, A. (1982). *Brief manual for the Differential Personality Questionnaire*. Unpublished manuscript, University of Minnesota.

Tellegen, A. (in press) *Manual for the Multidimensional Personality Questionnaire*. Minneapolis: University of Minnesota Press.

Thissen, D. (2003). MULTILOG [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Walton, K. E. (2005). *Using Item Response Theory to bridge the measurement gap between normal personality and psychopathy*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.

Widiger, T. A., & Lynam, D. R. (1998). Psychopathy and the Five-Factor Model of Personality. In T. Millon, E. Simonsen, M. Birket-Smith, & R. D. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behavior* (pp. 171–187). New York: Guilford Press.