Personality Disorders: Theory, Research, and Treatment
2013, Vol. 4, No. 1, 15–22

© 2012 American Psychological Association
1949-2715/12/$12.00 DOI: 10.1037/a0027948

An Item Response Theory Analysis of the *DSM–IV* Borderline Personality Disorder Criteria in a Population-Based Sample of 11- to 12-Year-Old Children

Jared D. Michonski
Dialectical Behavior Therapy Center of Seattle, Seattle, Washington

Carla Sharp and Lynne Steinberg
University of Houston

Mary C. Zanarini
McLean Hospital, Belmont, Massachusetts

Although a growing body of empirical literature provides some support for the diagnosis of borderline personality disorder (BPD) in youth, little is known about the internal structure of BPD and the performance of the individual diagnostic criteria, especially in younger samples. We used item response theory (IRT) methods to investigate the psychometric properties of the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (*DSM–IV*) BPD criteria in a large, population-based sample (n = 6,339) of young adolescents from the United Kingdom (ages 11 to 12). BPD was assessed using the Childhood Interview for *DSM–IV* Borderline Personality Disorder (CI-BPD; Zanarini, Horwood, Waylen, & Wolke, 2004). A single underlying dimension adequately accounted for covariation among the BPD criteria. Each criterion was found to be discriminating to a degree comparable to what has been reported in adult studies. BPD criteria were most informative within a range of severity of BPD pathology between +1 and +3 standard units. Five criteria were found to exhibit differential item functioning (DIF) between boys and girls. However, DIF balanced out for the total interview score. Despite the controversy associated with applying the borderline construct to youth, the current findings provide psychometric evidence in favor of doing so.

Keywords: borderline personality disorder, item response theory, differential item functioning, gender, Avon Longitudinal Study of Parents and Children

Beginning with the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*; American Psychiatric Association, 1994), application of the diagnosis of borderline personality disorder (BPD) to youth was permitted. Despite this allowance, diagnosing youth with BPD has engendered a great deal of reluctance. Several explanations for this reluctance have been offered. Clinicians may hesitate to provide personality disorder (PD) diagnoses because the PD label connotes severity and nonmalleability, which may negatively affect a developing child's self-concept or bias others' perceptions of the child (Kernberg, Weiner, & Bardenstein, 2000). Terr and Kernberg (1990) questioned diagnosing BPD before the onset of puberty and before the completion of identity formation. Many mental health professionals regard personality as lacking cohesiveness and stability before age 18 (Crick, Murray-Close, & Woods, 2005). Others have suggested that borderline features may occur as part of the normal developmental trajectory of adolescence (Meijer, Goedhart, & Treffers, 1998; Miller, Muehlenkamp, & Jacobson, 2008).

Despite these concerns, consideration of the borderline construct in children and adolescents is necessary if a better understanding of the development and etiology of BPD is to occur (Crick et al., 2005; Sharp & Bleiberg, 2007) and if early identification and prevention of BPD are to become a reality (Chanen, Jovev, McCutcheon, Jackson, & McGorry, 2008). Notably, a growing body of research has examined the viability of the *DSM* definition of BPD in youth. For instance, BPD can be reliably diagnosed in adolescents, appears to occur at similar rates across adolescent and adult inpatient settings, and the criteria have shown a degree of cohesiveness that is consistent with adult findings. Concurrent and, to a lesser extent, predictive validity have been demonstrated in

This article was published Online First May 28, 2012.

Jared D. Michonski, Dialectical Behavior Therapy Center of Seattle, Seattle, Washington; Carla Sharp and Lynne Steinberg, Department of Psychology, University of Houston; Mary C. Zanarini, McLean Hospital, Belmont, Massachusetts.

The authors are extremely grateful to all of the families who took part in this study; the midwives for their help in recruiting them; and the whole ALSPAC team, including the interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The U.K. Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provided core support for ALSPAC. This publication is the work of the authors, and Carla Sharp will serve as guarantor for the contents of this paper. This research was partially funded by a National Institute of Mental Health grant (R01 MH078757), a National Alliance for Research on Schizophrenia and Depression Young Investigators Award, and foundation funding from the Child and Family Program of the Menninger Clinic awarded to Carla Sharp.

Correspondence concerning this article should be addressed to Carla Sharp, Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204. E-mail: csharp2@uh.edu

several studies (for reviews see Miller et al., 2008; Sharp & Romero, 2007).

One approach to investigating the internal validity of BPD in youth that has yet to be used is item response theory (IRT). In recent years, IRT has gained increasing application in clinical psychology and psychiatry as a useful tool for evaluating the validity and utility of *DSM* criteria (e.g., Aggen, Neale, & Kendler, 2005). IRT constitutes a latent trait approach to psychological measurement, modeling the probability of endorsing an item (for dichotomous items) or response category (for polytomous items) as a function of an individual's standing on the underlying latent trait. For a full description of the general advantages of IRT over classical test theory approaches, see Embretson and Reise (2000), and for a discussion of the application of IRT in the context of child psychopathology assessment, see Sharp, Goodyer, and Croudace (2006).

Applied to diagnostic assessment, IRT addresses several key aspects of criterion functioning. First, IRT can be used to evaluate how well each criterion discriminates individuals in their standing along the continuum of disorder liability. Are all of the criteria discriminating, or do certain criteria perform poorly as indicators of disorder liability? Second, IRT can identify where along the latent continuum the threshold of endorsement for each criterion is located. Do the criteria discriminate in the same region of the liability continuum, or are certain criteria more difficult to endorse than others? Lastly, IRT can be used to evaluate group differences in the performance of the BPD criteria. Differential item functioning (DIF) exists when the relation of an item to the latent trait is different across population subgroups, such as gender. DIF occurs when individuals who have the same standing on the latent trait do not have the same probability of item endorsement.

Most applications of IRT assume unidimensionality and local independence; that is, covariation among items can be accounted for by a single common latent factor. To date, only three studies have examined the factor structure of the *DSM* criteria for BPD using youth samples (Becker, McGlashan, & Grillo, 2006; Chabrol et al., 2002; Sharp, Ha, Michonski, Venta, & Carbone, in press). Becker et al. (2006) performed principal component analysis using varimax factor rotation on the *DSM–III–R* symptom criteria. They regarded a four-factor solution as offering the most conceptual appeal. Conducting confirmatory factor analysis (CFA) on the *DSM–IV* criteria, Chabrol et al. (2002) favored a single-factor solution over alternative two- and three-factor models. However, these results should be viewed with some caution because both studies were limited by small sample size and did not use statistical procedures appropriate for ordinal data. In a much larger sample of inpatient adolescents ($N = 245$), Sharp, Ha, Michonski, Venta, and Carbone (in press) established a single-factor structure for *DSM–IV* borderline criteria. Additional support for unidimensionality comes from the adult BPD literature, in which a single-factor solution has been the predominant and most parsimonious finding (Aggen, Neale, Røysamb, Reichborn-Kjennerud, & Kendler, 2009; Clifton & Pilkonis, 2007; Feske et al., 2007; Fossati et al., 1999; Johansen, Karterud, Pedersen, Gude, & Falkum, 2004). Although a few studies of adult BPD have favored multidimensional solutions (e.g., Clarkin, Hull, & Hurt, 1993; Rosenberger & Miller, 1989; Sanislow, Grilo, & McGlashan, 2000; Sanislow et al., 2002), most of these were conducted with *DSM–III* or *DSM-III-R* criteria (i.e., eight instead of nine criteria). Only Sanislow et al. (2002) used the *DSM–IV*. They favored a three-factor solution; however, a single-factor model was also found to exhibit adequate fit and was arguably the better model, both in terms of model parsimony (i.e., factor correlations exceeded .90 for the three-factor solution) and in terms of conceptual appeal (e.g., abandonment fears loaded on the "affective dysregulation" factor along with affective instability and uncontrolled anger, rather than the "disturbed relatedness" factor). Thus, it seems reasonable to expect that the BPD criteria may adhere sufficiently to unidimensional factor structure in youth. However, this question remains to be tested in a younger, preadolescent sample.

A few studies have applied IRT methods to examine the BPD criteria in adults (e.g., Aggen et al., 2009; Feske et al., 2007; Jane, Oltmanns, South, & Turkheimer, 2007). In Feske et al. (2007) and Aggen et al. (2009), the BPD criteria were found to adequately discriminate adults in their standing along a continuum of BPD liability, and a high degree of BPD liability was required for each criterion to be rated as present. However, whether these findings hold for youth remains to be seen. Perhaps, as argued by Becker and colleagues (2006), BPD is a more diffuse pathology in youth samples, resulting in little internal consistency.

Studies examining the BPD criteria in adults have also demonstrated evidence of DIF as a function of gender (Aggen et al., 2009; cf., Jane et al., 2007). However, whether DIF is present in youth samples is unknown. Studies with community (Bernstein et al., 1993; Chabrol, Montovany, Chouicha, Callahan, & Mullet, 2001) and psychiatric samples (Grilo et al., 1996) have reported higher frequency of BPD among adolescent girls than boys, but the extent to which such differences reflect true gender differences versus DIF remains untested.

Against this background, the overall objective of the present study was to evaluate the performance of the *DSM* criteria for BPD in youth using IRT in a large, population-based sample of English children aged 11 to 12. First, because undimensional IRT assumes that the covariation among the BPD criteria can be accounted for by a single dimension, we evaluated whether a single factor underlies the criteria. Second, the utility of each individual BPD criterion was evaluated on the basis of IRT discrimination and threshold parameters. Finally, the presence of DIF across gender was evaluated. In anticipation of DIF, the evaluation of dimensionality was conducted separately for girls and boys.

Method

Participants

The sample consisted of children who participated in the Avon Longitudinal Study of Parents and Children (ALSPAC), a prospective birth cohort study intended to be representative of Great Britain as a whole and designed to identify how an individual's genotype and environment affect health and development. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Study aims have been described previously (Golding et al., 2001). In brief, 14,541 pregnant women residing in Bristol, England, with an expected delivery date between April 1991 and December 1992 were enrolled. Data used in the present investigation were collected during a clinic visit occurring between January 2003 and January 2005, when children were approximately 11 years old

(Focus 11+). The clinic visit involved a range of physiological and psychological measures, including a British version of the Childhood Interview for *DSM–IV* Borderline Personality Disorder (CI-BPD; Zanarini, Horwood, Waylen, & Wolke, 2004). Of the 7,149 children who attended the Focus 11+ clinic visit, 6,423 were administered the CI-BPD. Of these, 6,409 children received ratings for at least one of the nine CI-BPD items. However, 70 twin pairs were included among these participants. To eliminate data analytic problems associated with dependency, a single child from each twin pair was randomly excluded from analysis, resulting in a final sample size of 6,339. Of these, 3,071 (48.45%) were male and 3,268 (51.55%) were female. Forty-two participants met study criteria for BPD, including 19 male (45.2%) and 23 female (54.8%) children. Ninety-six percent of children were White (n = 6,205). The mean age of children at the Focus 11+ clinic visit was 140.97 months (SD = 3.86), or 11.75 years old.

Measure

Study participants were administered the British version of the CI-BPD (Zanarini et al., 2004), a semistructured interview that assesses *DSM–IV* BPD in children and adolescents; all nine of the *DSM–IV* BPD criteria were included. It is based on the borderline module of the Diagnostic Interview for *DSM–IV* Personality Disorders (Zanarini, Frankenburg, Sickel, & Yong, 1996). The U.K. version of the CI-BPD was modified slightly for the British sample (e.g., saying "cross" instead of "angry"). Criterion ratings are made after asking a series of questions associated with each criterion. Each criterion is rated with a score of 0 (*absent*), 1 (*probably present*), or 2 (*definitely present*). A diagnosis is assigned if the child receives a rating of 2 on five or more of the diagnostic criteria. Beyond the original validation study of the American version of the CI-BPD (Zanarini, 2003), the psychometric properties of the CI-BPD were recently examined in an inpatient sample of adolescents (N = 245; Sharp et al., in press). Internal consistency (Cronbach's α = .80), interrater reliability (κ = .89), convergent validity (with two questionnaire-based measures of BPD), and concurrent validity (with Axis I psychopathology and deliberate self-harm) were excellent. Although CI-BPD diagnosis was only moderately related to clinician diagnosis at discharge in this study (κ = .34. p < .001), and in another recent study using the CI-BPD (κ = .47, p = .001; Chang, Sharp, & Ha, 2011), this may be explained by the fact that the CI-BPD was administered in both studies at admission, whereas clinician diagnosis was reported at discharge due to a reluctance on the part of the clinicians to diagnose PD at admission.

Consistent with other studies (Chang et al., 2011; Sharp, Mosko, Chang, & Ha, 2011), the CI-BPD showed good internal reliability in the present sample (Cronbach's α of .78). Item-total correlations ranged from .46 (suicidal gestures) to .73 (uncontrolled anger), and interitem correlations ranged from .16 (abandonment fears with impulsivity) to .48 (uncontrolled anger with affective instability).

Data Analytic Plan

The IRT model fitting and the computation of the test statistics were performed using a beta version of IRTPRO (Cai, du Toit, & Thissen, 2011; Thissen, 2009). Goodness of fit of the models was evaluated using the M_2 statistics and its associated root mean square error of approximation (RMSEA) value (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Maydeu-Olivares & Joe, 2005, 2006; Thissen, 2009), as well as the standardized local dependence (LD) χ^2 indices (based on the LD index proposed by Chen & Thissen, 1997). LD indicates that the observed covariation among responses to the items in an item pair exceeds that predicted by the model. The LD indices are standardized χ^2 values; values 10 or greater are considered noteworthy (Thissen, 2009) and thus challenge the assumption of unidimensionality.

The graded-response IRT model (Samejima, 1969) was fitted to the nine BPD criteria. The graded-response model is useful for polytomously scored items such as those of the CI-BPD. The graded-response model is an extension of the 2 parameter logistic (2PL) model and divides the response process into binary pieces representing the probability of scoring in or above a given response category for each item as a function of the underlying construct. For each item, two types of parameters are estimated—*discrimination* (or slope) and *thresholds* (Embretson & Reise, 2000). The discrimination parameter represents the degree of association between the item response and the underlying construct. For a three-category item, scored 0, 1, or 2, the two estimated thresholds reflect the level of latent BPD liability needed to score above 1 (*probably present*) and 2 (*definitely present*), respectively, on the CI-BPD items with .50 probability. The probability of a response in a particular category k (i.e., 0, 1, or 2) is given by the probability of observing k or higher minus the probability of observing $k + 1$ or higher.

The presence of DIF was investigated using the approach advanced by Thissen, Steinberg, and Wainer (1993). This approach simultaneously calibrates item parameter estimates across groups, in this case gender. Differences in parameter estimates between groups are evaluated using comparison tests. To implement this approach, a subset of items ("anchor items") is identified as a means to "link" the groups (allowing for an estimated group mean difference in the underlying construct). Edelen, Thissen, Teresi, Kleinman, and Ocepek-Welirkson (2006) recommend identifying anchor items by using an exploratory, iterative process in which each item is initially tested for DIF by using all other items as the anchor set. Items not showing DIF at this step are regarded as anchor items; the remaining items, referred to as the "studied items," are then evaluated for DIF. Wald tests based on the procedure proposed by Lord (1977), providing separate χ^2 statistics for the discrimination and threshold parameters for each studied item, are used to evaluate the presence of DIF. When DIF is detected, effect sizes for the threshold and/or slope parameters will aid the description and interpretation of the group differences (Steinberg & Thissen, 2006).

Results

IRT Analyses

Unidimensional IRT model. In separate analyses of the item response data for girls and boys, the graded unidimensional IRT model showed satisfactory fit (girls: M_2 (135) = 290.87, p < .001; $RMSEA$ = 0.02; boys: M_2 (135) = 301.26, p < .001; $RMSEA$ = 0.02). None of the standardized χ^2 indices of LD approached the value of 10.0 for boys or girls. For boys, the largest LD value was observed between the uncontrolled anger and impulsivity criteria

(*LD* $\chi^2 = 5.3$), whereas, for girls, the highest value was observed between the uncontrolled anger and unstable relationships criteria (*LD* $\chi^2 = 3.1$). Taken together, the findings of unidimensionality and local independence offered justification for proceeding with unidimensional IRT analyses.

The graded model slope parameter estimates showed that all symptom criteria were found to be adequately discriminating for boys and girls. The slope parameters are analogous to factor loadings in traditional or CFA; in fact, slope parameters can be translated into factor loadings (e.g., see McLeod, Swygert, & Thissen, 2001, p. 199). Slope parameters that are 1.0 or greater are considered substantial. The discrimination (slope) parameters ranged from 1.44 (impulsivity) to 2.44 (paranoid ideation) for boys and from 1.28 (impulsivity) to 2.92 (identity disturbance) for girls.

Threshold parameters for boys and girls were all located above the mean. This is to be expected given that the CI-BPD is a clinical measure that was administered to a community sample of children. For boys, thresholds corresponding to a rating of 1 (*probably present*) ranged from 0.80 (uncontrolled anger) to 2.40 (abandonment fears), and thresholds corresponding to a rating of 2 (*definitely present*) ranged from 1.47 (impulsivity) to 3.86 (abandonment fears). For girls, thresholds for a response rating of 1 ranged from 0.94 (uncontrolled anger) to 2.48 (suicidal behaviors), and thresholds for a response rating of 2 ranged from 2.00 (uncontrolled anger/impulsivity) to 3.06 (suicidal behaviors). Generally speaking, symptoms dealing with emotional reactivity or poor impulse control (i.e., uncontrolled anger, affective instability, and impulsivity) were the easier to endorse compared with suicidal behaviors and abandonment fears that were the more “difficult” to endorse.

Detection of DIF. The first step in conducting the IRT analyses was to identify a set of anchor items for linking the boy and girl subgroups. To do so, each item was initially tested for DIF using all of the other items as a tentative anchor. Three items emerged as not exhibiting DIF as evidenced by nonsignificant Wald test (χ^2) statistics ($p > .05$): emptiness, identity disturbance, and paranoid ideation. However, because of similarity of item parameter estimates, an additional item (affective instability) warranted consideration as a potential anchor item although not identified according to the conservative guideline suggested by Edelen et al. (2006) ($p = .05$). Therefore, affective instability was included in the next step for identifying the anchor set. Evaluating these four items for DIF, excluding the other items, none of the Wald statistics approached significance (all *p* values exceeded .35), indicating that all four items can be included in the final anchor set. The remaining five items constituted the studied items in the subsequent DIF analyses.

For evaluating the Wald tests for the five studied items, type I error rate was controlled using the Benjamini-Hochberg (B-H) multiple comparisons procedure (Benjamini & Hochberg, 1995). All of the studied items exhibited DIF (Table 1). Three of the items (uncontrolled anger, suicidal behaviors, and impulsivity) showed DIF concentrated in the threshold (*b*) parameters, whereas the other two items (abandonment fears and unstable relationships) exhibited DIF with respect to the discrimination (*a*) and threshold (*b*) parameters. The final calibration of item parameters was performed by fitting a model in which item parameters found to exhibit DIF were estimated separately for boys and girls, whereas item parameters for the anchor items and for those showing nonsignificant differences were constrained to be equal across gender (see Table 2). No significant difference in the latent trait means of BPD liability was detected; the mean level of BPD liability for girls was 0.05 standard units higher than boys.

Threshold DIF. Uncontrolled anger, suicidal behaviors, and impulsivity showed DIF concentrated in threshold parameters. For all three items, the direction of DIF was such that it was “easier” for boys than for girls to be rated as exhibiting the symptom at the same level of BPD liability. For uncontrolled anger, the difference between boys and girls was 0.19 standard units for the first threshold and 0.28 for the second threshold. Regarding the latter threshold, this means that girls were over one quarter a standard unit higher than boys in the value of BPD liability necessary to have a 50–50 chance of being rated as “definitely” exhibiting uncontrolled anger. The effect sizes for threshold differences for suicidal behaviors were also rather small: 0.34 for the first threshold and 0.20 for the second threshold. However, the effect sizes for threshold differences for impulsivity were more substantial: 0.58 and 0.46 for the first and second thresholds, respectively. In other

Table 1
Wald Statistics for Model Using Affective Instability, Emptiness, Identity Disturbance, and Paranoid Ideation as Anchor Items

Item	χ^2	Observed *p* value	B-H Critical *p* value	Rank
Threshold DIF				
Impulsivity	**78.0**	**0.0001**	**0.0100**	**1**
Uncontrolled anger	**19.2**	**0.0001**	**0.0200**	**2**
Suicidal behaviors	**16.7**	**0.0002**	**0.0300**	**3**
Unstable relationships	**16.1**	**0.0003**	**0.0400**	**4**
Abandonment fears	**13.9**	**0.0009**	**0.0500**	**5**
Discrimination DIF				
Abandonment fears	**7.3**	**0.0070**	**0.0100**	**1**
Unstable relationships	**6.1**	**0.0133**	**0.0200**	**2**
Impulsivity	2.4	0.1256	0.0300	3
Uncontrolled anger	0.0	0.8246	0.0400	4
Suicidal behaviors	0.0	0.8436	0.0500	5

Note. Values in bold are statistically significant. Observed *p* values are based on 1 and 2 degrees of freedom for discrimination (*a*) and threshold (*b*) parameters, respectively.

Table 2
IRT Parameter Estimates for Model in Which Parameters Not Showing DIF Are Constrained to Equality Across Gender

Item	Gender	a	b_1	b_2
Anchor items				
Affective instability	Both	2.28 (.09)	1.07 (.03)	2.13 (.06)
Emptiness	Both	2.27 (.11)	1.79 (.05)	2.62 (.08)
Identity disturbance	Both	2.65 (.13)	1.63 (.04)	2.51 (.07)
Paranoid ideation	Both	2.39 (.10)	1.42 (.03)	2.46 (.07)
Threshold DIF				
Uncontrolled anger	Boys	2.37 (.08)	0.80 (.03)	1.79 (.06)
	Girls	2.37 (.08)	0.99 (.04)	2.07 (.07)
Suicidal behaviors	Boys	1.83 (.10)	2.24 (.11)	2.97 (.18)
	Girls	1.83 (.10)	2.58 (.14)	3.17 (.19)
Impulsivity	Boys	1.36 (.18)	0.95 (.08)	1.52 (.14)
	Girls	1.36 (.18)	1.53 (.13)	1.98 (.18)
Discrimination and threshold DIF				
Abandonment fears	Boys	1.56 (.12)	2.38 (.12)	3.84 (.24)
	Girls	2.10 (.13)	1.85 (.07)	2.82 (.12)
Unstable relationships	Boys	2.16 (.13)	1.57 (.05)	2.55 (.10)
	Girls	1.75 (.10)	1.44 (.05)	2.70 (.12)

Note. Values enclosed in parentheses represent standard errors of the parameter estimates. a represents the discrimination parameter; b_1 and b_2 represent the first and second threshold parameters, respectively.

words, considering the first threshold, girls were over one half a standard unit higher than boys in the level of BPD liability necessary to have a 50–50 chance of being rated as "probably" positive (or higher) for suicidal behavior.

Discrimination DIF. Abandonment fears and unstable relationships showed significant DIF in discrimination across gender. The nature of DIF for abandonment fears was such that the item was more discriminating for girls (a = 2.10) than for boys (a = 1.56). Framed in terms of effect size, the relationship between abandonment fears and BPD liability was 1.35 times stronger for girls than boys. DIF was in the opposite direction for unstable relationships: The item was more discriminating for boys (a = 2.16) than for girls (a = 1.75). The relationship between unstable relationships and BPD liability was 1.23 times stronger for boys than girls.

Interpreting DIF for these items is complicated by the fact that the threshold parameters (b) also differed significantly across gender. When this is the case, the effect is inherently multivariate (Steinberg & Thissen, 2006). As depicted in Figure 1, the trace lines differ in their slopes as well as their right-left locations (i.e., threshold parameters) for boys and girls. Steinberg and Thissen (2006) recommended plotting expected item scores as a function of their underlying latent trait to reduce the complexity of interpretation. The expected score curves for abandonment fears (see Figure 1) show that girls were more likely to be rated as having abandonment fears at lower levels of the underlying trait relative to boys. For example, for girls the expected score approached a rating of 1 (*probably present*) at a level of BPD liability that is approximately 2.3 standard units above the mean, whereas for boys the expected score did not approach 1 until a level of BPD liability that is approximately 3 standard units above the mean. The discrepancy (although outside of the range of Figure 1) is even greater as the expected score approaches a rating of 2 (*definitely present*), as is evident in the greater rate of change (slope) for girls relative to boys.

For unstable relationships, DIF appears to be of less practical impact (see Figure 1). The expected score approaches a rating of 1 (*probably present*) at similar levels of BPD liability for boys and girls—at approximately 2.0 standard units above the mean. However, DIF is somewhat more pronounced (although not visible in Figure 1) as the expected score approaches a rating of 2 (*definitely present*). The expected score for boys approaches 2 at a lower level of BPD liability than it does for girls.

Overall test curves. The impact of DIF on the test as a whole can be evaluated by considering the test characteristic curve. The test characteristic curve models the expected test score (i.e., the expected sum of the nine BPD criterion scores) as a function of

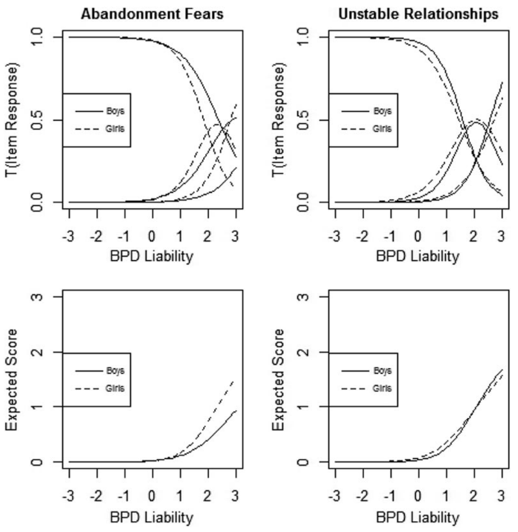

Figure 1. Category response curves (upper) and expected score curves (lower) for abandonment fears and unstable relationships, respectively.

each value on the underlying construct continuum. Figure 2 (top) shows that the effect of DIF is cancelled out at the level of the test (interview). That is, the expected interview score is identical for boys and girls at any given level of the underlying construct as indicated by the virtually overlapping curves. Figure 2 (bottom) displays the test information curves for both genders for the CI-BPD interview as a whole. These curves indicate where along the continuum of the underlying construct the measurement is most precise. As depicted in Figure 2, the BPD score is most informative at the positive end of the continuum, primarily within the range of 1.0–3.0.

Discussion

The purpose of the present study was to evaluate the performance of the *DSM* criteria in a population-based sample of 11- to 12-year-old youth using IRT methods by evaluating (a) the underlying factor structure of the BPD criteria, (b) the utility of each individual BPD criterion on the basis of IRT discrimination and threshold parameters, and (c) the measurement equivalence of each criterion (DIF) across gender.

A unidimensional model was found to fit the BPD criteria well. This finding is consistent with CFA studies from the adult literature (Aggen et al., 2009; Clifton & Pilkonis, 2007; Feske et al., 2007; Fossati et al., 1999; Johansen et al., 2004; Sanislow et al., 2002), as well as one study using an adolescent sample (Chabrol et al., 2002), and indicates that the *DSM* criteria constitute a coherent combination of traits and symptoms, even in pre- and young-adolescent youth. This is consistent with the growing trend to view psychiatric diagnoses, especially PDs, as continuously distributed phenomena rather than discrete categories (e.g., Widiger & Samuel, 2005). A dimensional perspective may be particularly important for conceptualizing BPD pathology among youth because it is better able to account for the developmental fluctuations and increased heterogeneity that have been reported in younger samples (Miller et al., 2008). Further, that the criteria conformed to a unidimensional model is notable in that the BPD criteria were selected via expert consensus, with limited reliance on psychometric theory (Aggen et al., 2009).

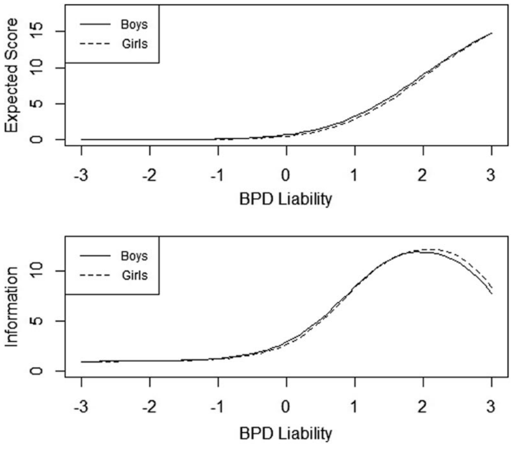

Figure 2. Test characteristic curves (upper) and test information curves (lower).

The IRT analyses produced several key findings. All nine criteria were found to be discriminating for boys and girls. The criteria were most discriminating at the high (positive) end of the BPD liability continuum, with measurement precision (information) for the instrument as a whole the highest between +1 and +3 on the underlying construct continuum. This finding is not surprising given that the CI-BPD, an interview designed to identify clinical cases of borderline personality, was given in a community sample. Aggen et al. (2009) reported similar findings in their general population sample.

Regarding specific criteria, the threshold parameters indicate that abandonment fears and suicidal behavior were the most "difficult" for boys and girls in that they required the highest level of BPD liability to be expected to be rated as present. These findings are consistent with studies showing that abandonment fears are the least commonly exhibited symptom in adult patient samples (Becker, Grilo, Edell, & McGlashan, 2002; Clifton & Pilkonis, 2007; McGlashan et al., 2005) and that suicidal behaviors are rare in young children (e.g., Resch, Parzer, & Brunner, 2008).

A final key finding to emerge was that several BPD criteria functioned differently across gender. Three items showed DIF with respect to threshold parameters (uncontrolled anger, suicidal behaviors, and impulsivity), and two items showed DIF with respect to discrimination and threshold parameters (abandonment fears and unstable relationships). These instances of DIF may have resulted from the wording of a given interview item such that it favors members of a particular subgroup. In this regard, the current findings suggest the need for possible modifications when assessing for the presence of certain criteria, although such modifications may be largely specific to use of the CI-BPD. Interviewers may be encouraged to inquire about gender-relevant manifestations of certain symptoms. For example, regarding the impulsivity criterion, when assessing young females, it may be important to ask about nonaggressive and nonovert aggressive forms of impulsive behavior (Crick & Grotpeter, 1995; Crick & Werner, 1998; Loeber, Burke, Lahey, Winters, & Zera, 2000). The same may also hold true for the uncontrolled anger criterion: Examiners should consider inquiring about less overt forms of anger expression (Crick & Grotpeter, 1995; Crick & Werner, 1998). It may also be the case that DIF occurred for the anger criterion because the children and/or study interviewers tended to regard displays of anger as more socially acceptable in boys than in girls (Cole, Teti, & Zahn-Waxler, 2003; Underwood, Galen, & Paquette, 2001; Zhou, Eisenberg, Wang, & Resier, 2004). On the other hand, DIF may have occurred because, in addition to the common factor that is being measured, a given item also taps a specific factor that really does differ across subgroups (Wicherts & Dolan, 2010). For example, in addition to measuring "BPD liability" (common factor), the abandonment fears criterion may also reflect a particular orientation toward relationships (specific factor), such as one characterized by a preference for dyadic relationships and deeper emotional connection. Indeed, such a relational style has been observed to be more common in girls than boys (Rose & Rudolph, 2006). The same may also hold true for impulsivity, which would be consistent with findings that, among children with attention deficit hyperactivity disorder, boys exhibit greater degrees of hyperactivity, impulsivity, and externalizing problems (Gershon, 2002). However, it should be noted that these explanations are speculative and require further investigation. More important, that

DIF balances out at the level of the entire set of criteria implies that users of the CI-BPD should include all nine criteria to avoid the consequences of DIF on BPD scores.

Taken together, the current study adds to a growing body of research extending the borderline construct to youth (Miller et al., 2008). In fact, this study provides evidence for extending the construct even further—to pre- and young-adolescent children, where the empirical base is even thinner. The present study offers several methodological advantages over previous investigations evaluating the BPD criteria in youth. A large, nationally representative sample was used, and statistical procedures appropriate for analyzing categorical data were used. Moreover, this was the first study to evaluate the measurement equivalence of the BPD criteria across gender in a youth sample.

However, several limitations also deserve note. First, the present study addressed only internal aspects of validity; research connecting CI-BPD scores to other outcomes would add to our understanding of BPD in children. Second, our findings do not speak to the stability or longitudinal course of BPD pathology. BPD symptomatology in children as assessed by the *DSM* criteria may represent a time-limited manifestation of a different underlying psychopathology, and more research is needed to illuminate the developmental trajectory of BPD pathology.

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: A Journal of Research in Psychiatry and the Allied Sciences, 35,* 475–487. doi:10.1017/S0033291704003563

Aggen, S. H., Neale, M. C., Røysamb, E., Reichborn-Kjennerud, T., & Kendler, K. S. (2009). A psychometric evaluation of the DSM-IV borderline personality disorder criteria: Age and sex moderation of criterion functioning. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 39,* 1967–1978. doi:10.1017/S0033291709005807

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

Becker, D. F., Grilo, C. M., Edell, W. S., & McGlashan, T. H. (2002). Diagnostic efficiency of borderline personality disorder criteria in hospitalized adolescents: Comparison with hospitalized adults. *The American Journal of Psychiatry, 159,* 2042–2047. doi:10.1176/appi.ajp.159.12.2042

Becker, D. F., McGlashan, T. H., & Grilo, C. M. (2006). Exploratory factor analysis of borderline personality disorder criteria in hospitalized adolescents. *Comprehensive Psychiatry, 47,* 99–105.; doi:10.1016/j.comppsych.2005.07.003

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 57,* 289–300.

Bernstein, D. P., Cohen, P., Velez, C. N., Schwab-Stone, M., Siever, L. J., & Shinsato, L. (1993). Prevalence and stability of the DSM-III-R personality disorders in a community-based survey of adolescents. *The American Journal of Psychiatry, 150,* 1237–1243.

Cai, L., du Toit, S. H. C., & Thissen, D. (2011). IRTPRO: Flexible professional item response theory modeling for patient reported outcomes [Computer software]. Chicago, IL: SSI International.

Cai, L., Maydeu-Olivares, A., Coffman, D. L., & Thissen, D. (2006). Limited information goodness-of-fit testing of item response theory models for sparse 2^p tables. *British Journal of Mathematical and Statistical Psychology, 59,* 173–194. doi:10.1348/000711005X66419

Chabrol, H., Chouicha, K., Montovany, A., Callahan, S., Duconge, E., & Sztulman, H. (2002). Personality disorders in a nonclinical sample of adolescents. *L'Encephale: Revue de Psychiatrie Clinique Biologique et Thérapeutique, 28,* 520–524.

Chabrol, H., Montovany, A., Chouicha, K., Callahan, S., & Mullet, E. (2001). Frequency of borderline personality disorder in a sample of French high school students. *The Canadian Journal of Psychiatry, 46,* 847–849.

Chanen, A. M., Jovev, M., McCutcheon, L. K., Jackson, H. J., & McGorry, P. D. (2008). Borderline personality disorder in young people and the prospect for prevention and early intervention. *Current Psychiatry Reviews, 4,* 48–57. doi:10.2174/157340008783743820

Chang, B., Sharp, C., & Ha, C. (2011). The criterion validity of the Borderline Personality Disorder Features Scale for Children. *Journal of Personality Disorders, 25,* 492–503.

Chen, W. H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22,* 265–289.

Clarkin, J. F., Hull, J. W., & Hurt, S. W. (1993). Factor structure of borderline personality disorder criteria. *Journal of Personality Disorders, 7,* 137–143. doi:10.1521/pedi.1993.7.2.137

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. doi:10.1016/j.comppsych.2006.07.002

Cole, P. M., Teti, L. O., & Zahn-Waxler, C. (2003). Mutual emotion regulation and the stability of conduct problems between preschool and early age. *Development and Psychopathology, 15,* 1–18. doi:10.1017/S0954579403000014

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710–722. doi:10.2307/1131945

Crick, N. R., Murray-Close, D., & Woods, K. (2005). Borderline personality features in childhood: A short-term longitudinal study. *Development and Psychopathology, 17,* 1051–1070. doi:10.1017/S0954579405050492

Crick, N. R., & Werner, N. E. (1998). Response decision processes in relational and overt aggression. *Child Development, 69,* 1630–1639.

Edelen, M. O., Thissen, D., Teresi, J., Kleinman, M., & Ocepek-Welirkson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-Mental Status Examination. *Medical Care, 44,* S134–S142. doi:10.1097/01.mlr.0000245251.83359.8c

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* London, UK: Lawrence Erlbaum.

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. doi:10.1521/pedi.2007.21.4.418

Fossati, A., Maffei, C., Bognato, M., Donati, D., Namia, C., & Novella, L. (1999). Latent structure analysis of DSM-IV borderline personality disorder criteria. *Comprehensive Psychiatry, 40,* 72–79. doi:10.1016/S0010-440X(99)90080-9

Gershon, J. (2002). A meta-analytic review of gender differences in ADHD. *Journal of Attention Disorders, 5,* 143–154. doi:10.1177/108705470200500302

Golding, J., Pembrey, M., Jones, R., & the ALSPAC Study Team (2001). ALSPAC—The Avon Longitudinal Study of Parents and Children: I. Study methodology. *Pediatric and Perinatal Epidemiology, 15,* 74–87. doi:10.1046/j.1365-3016.2001.00325.x

Grilo, C. M., Becker, D. F., Fehon, D. C., Walker, M. L., Edell, W. S., & McGlashan, T. H. (1996). Gender differences in personality disorders in psychiatrically hospitalized adolescents. *The American Journal of Psychiatry, 153,* 1089–1091.

Jane, J. S., Oltmanns, T. F., South, S. C., & Turkheimer, E. (2007). Gender bias in diagnostic criteria for personality disorders: An item response theory analysis. *Journal of Abnormal Psychology, 116,* 166–175. doi: 10.1037/0021-843X.116.1.166

Johansen, M., Karterud, S., Pedersen, G., Gude, T., & Falkum, E. (2004). An investigation of the prototype validity of the borderline DSM-IV construct. *Acta Psychiatrica Scandinavica, 109,* 289–298. doi:10.1046/j.1600-0447.2003.00268.x

Kernberg, P. F., Weiner, A. S., & Bardenstein, K. K. (2000). *Personality disorders in children and adolescents.* New York, NY: Basic Books.

Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorders: A review of the past 10 years, Part I. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 1468–1484. doi:10.1097/00004583-200012000-00007

Lord, F. M. (1977). A study of item bias, using item characteristic curve theory. In Y. H. Portinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19–29). Amsterdam, The Netherlands: Swets and Zeitlinger.

Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association, 100,* 1009–1020. doi:10.1198/016214504000002069

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71,* 713–732. doi:10.1007/s11336-005-1295-9

McGlashan, T. H., Grilo, C. M., Sanislow, C. A., Ralevski, E., Morey, L. C., Gunderson, J. G., . . . Pagano, M. (2005). Two-year prevalence and stability of individual criteria for schizotypal, borderline, avoidant, and obsessive-compulsive personality disorders: Toward a hybrid model of Axis II disorders. *The American Journal of Psychiatry, 162,* 883–889. doi:10.1176/appi.ajp.162.5.883

McLeod, L. D., Swygert, K., & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 189–216). Mahwah, NJ: Lawrence Erlbaum.

Meijer, M., Goedhart, A. W., & Treffers, P. D. (1998). The persistence of borderline personality disorder in adolescence. *Journal of Personality Disorders, 12,* 13–22. doi:10.1521/pedi.1998.12.1.13

Miller, A. L., Muehlenkamp, J. J., & Jacobson, C. M. (2008). Fact or fiction: Diagnosing borderline personality disorder in adolescents. *Clinical Psychology Review, 28,* 969–981. doi:10.1016/j.cpr.2008.02.004

Resch, F., Parzer, P., & Brunner, R. (2008). Self-mutilation and suicidal behavior in children and adolescents: Prevalence and psychosocial correlates: Results of the BELLA study. *European Child & Adolescent Psychiatry, 17,* 92–98. doi:10.1007/s00787-008-1010-3

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132,* 98–131.

Rosenberger, P. H., & Miller, G. A. (1989). Comparing borderline definitions: DSM-III borderline and schizotypal personality disorders. *Journal of Abnormal Psychology, 98,* 161–169. doi:10.1037/0021-843X.98.2.161

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph Supplement,* 34 (Monograph No. 17).

Sanislow, C. A., Grilo, C. M., & McGlashan, T. H. (2000). Factor analysis of the DSM-III-R borderline personality disorder in psychiatric inpatients. *The American Journal of Psychiatry, 157,* 1629–1633. doi: 10.1176/appi.ajp.157.10.1629

Sanislow, C. A., Grilo, C. M., Morey, L. C., Bender, D. S., Skodol, A. E., Gunderson, J. G., . . . McGlashan, T. H. (2002). Confirmatory factor analysis of DSM-IV criteria for borderline personality disorder: Findings from the Collaborative Longitudinal Personality Disorders Study. *American Journal of Psychiatry, 159,* 284–290. doi:10.1176/appi.ajp.159.2.284

Sharp, C., & Bleiberg, E. (2007). Borderline personality disorder in children and adolescents. In A. Martin and F. Volkmar (Eds.), *Lewis's child and adolescent psychiatry: Comprehensive textbook* (4th ed., pp. 680–691). Baltimore, MD: Lippincott Williams and Wilkins.

Sharp, C., Goodyer, I. M., & Croudace, T. J. (2006). The Short Mood and Feelings Questionnaire (SMFQ): A unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of Abnormal Child Psychology, 34,* 365–377. doi:10.1007/s10802-006-9027-x

Sharp, C., Ha, C., Michonski, J., Venta, A., & Carbone, C. (in press). Borderline Personality Disorder in adolescents: Evidence in support of the CI-BPD in a sample of adolescent inpatients. *Comprehensive Psychiatry.*

Sharp, C., Mosko, O., Chang, B., & Ha, C. (2011). The cross-informant concordance and concurrent validity of the Borderline Personality Features Scale for Children in a sample of male youth. *Clinical Child Psychology and Psychiatry, 16,* 335–349.

Sharp, C., & Romero, C. (2007). Borderline Personality Disorder: A comparison between children and adults. *Bulletin of the Menninger Clinic, 71,* 85–114. doi:10.1521/bumc.2007.71.2.85

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods, 11,* 402–415. doi:10.1037/1082-989X.11.4.402

Terr, L. C., & Kernberg, P. F. (1990). Debate forum—Resolved: Borderline personality disorder exists in children under twelve. *Journal of the American Academy of Child & Adolescent Psychiatry, 29,* 478–482. doi:10.1097/00004583-199005000-00025

Thissen, D. (2009). The MEDPRO project: An SBIR project for a comprehensive IRT and CAT software system—IRT software. In D. J. Weiss (Ed.), *Proceedings of the 2009 GMAC Conference on Computerized Adaptive Testing.* Retrieved from www.psych.umn.edu/ psylabs/CATCentral/

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models (pp. 67–113). In P. W. Holland & H. Wainer (Eds.), *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum.

Underwood, M. K., Galen, B. R., & Paquette, J. A. (2001). Top ten challenges for understanding gender and aggression in children: Why can't we all just get along? *Social Development, 10,* 248–266. doi: 10.1111/1467-9507.00162

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29,* 39–47. doi:10.1111/j.1745-3992.2010.00182.x

Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition. *Journal of Abnormal Psychology, 114,* 494–504. doi:10.1037/0021-843X.114.4.494

Zanarini, M. C. (2003). *Childhood Interview for DSM-IV Borderline Personality Disorder (CI-BPD).* Belmont, MA: McLean Hospital.

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

Zanarini, M. C., Horwood, J., Waylen, A., & Wolke, D. (2004). *The UK version of the Childhood Interview for DSM-IV Borderline Personality Disorder (UK-CI-BPD).* Bristol, UK: University of Bristol, Department of Community Medicine, Unit of Perinatal and Pediatric Epidemiology.

Zhou, Q., Eisenberg, N., Wang, Y., & Resier, M. (2004). Chinese children's effortful control and dispositional anger/frustration: Relations to parenting styles and children's social functioning. *Developmental Psychology, 40,* 352–366. doi:10.1037/0012-1649.40.3.352