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, & Trefers, 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
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 validity and utility of DSM criteria for BPD. 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 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 unidimensional 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 Child-hood 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 IRT analyses, 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, de Toit, & Thissen, 2011; Thissen, 2009). Goodness of fit of the models was evaluated using the M2 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) χ² 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 χ² 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 (Embreton & 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 + 1 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 χ² 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 χ² 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 ($\chi^2$) 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>
<thead>
<tr>
<th>Item</th>
<th>$\chi^2$</th>
<th>Observed $p$ value</th>
<th>B-H Critical $p$ value</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold DIF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulsivity</td>
<td>78.0</td>
<td>0.0001</td>
<td>0.0100</td>
<td>1</td>
</tr>
<tr>
<td>Uncontrolled anger</td>
<td>19.2</td>
<td>0.0001</td>
<td>0.0200</td>
<td>2</td>
</tr>
<tr>
<td>Suicidal behaviors</td>
<td>16.7</td>
<td>0.0002</td>
<td>0.0300</td>
<td>3</td>
</tr>
<tr>
<td>Unstable relationships</td>
<td>16.1</td>
<td>0.0003</td>
<td>0.0400</td>
<td>4</td>
</tr>
<tr>
<td>Abandonment fears</td>
<td>13.9</td>
<td>0.0009</td>
<td>0.0500</td>
<td>5</td>
</tr>
<tr>
<td>Discrimination DIF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abandonment fears</td>
<td>7.3</td>
<td>0.0070</td>
<td>0.0100</td>
<td>1</td>
</tr>
<tr>
<td>Unstable relationships</td>
<td>6.1</td>
<td>0.0133</td>
<td>0.0200</td>
<td>2</td>
</tr>
<tr>
<td>Impulsivity</td>
<td>2.4</td>
<td>0.1256</td>
<td>0.0300</td>
<td>3</td>
</tr>
<tr>
<td>Uncontrolled anger</td>
<td>0.0</td>
<td>0.8246</td>
<td>0.0400</td>
<td>4</td>
</tr>
<tr>
<td>Suicidal behaviors</td>
<td>0.0</td>
<td>0.8436</td>
<td>0.0500</td>
<td>5</td>
</tr>
</tbody>
</table>

*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.
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 \( (a_{\text{girls}} = 2.10) \) than for boys \( (a_{\text{boys}} = 1.56) \). 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 \( (a_{\text{boys}} = 2.16) \) than for girls \( (a_{\text{girls}} = 1.75) \). 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 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.](image-url)
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).

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 & Grotpeeter, 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 & Grotpeeter, 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 (Wichert & 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


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.