An 11-Item Measure to Assess Borderline Traits in Adolescents: Refinement of the BPFSC Using IRT

Carla Sharp
University of Houston and The Menninger Clinic,
Houston, Texas

Lynne Steinberg
University of Houston

Jeff Temple
University of Texas Medical Branch Health

Elizabeth Newlin
The Menninger Clinic, Houston, Texas

Despite historical concerns about the validity of the construct of borderline personality disorder (BPD) in adolescence, there is now general consensus that BPD in adolescence constitutes a valid and reliable diagnosis. Yet the development and refinement of measures to assess borderline traits in adolescents is in its infancy. Moreover, brief and easy-to-administer measures of borderline traits for use in large-scale studies do not exist. The aim of the current study was to evaluate the Borderline Personality Features Scale for Children (BPFSC; Crick, Murray-Close, & Woods, 2005) and develop a short version of the BPFSC through the use of item response theory (IRT) methods. BPFSC data from a community sample of 964 adolescents (mean age = 15.1 years, SD = .79; 55.9% female) were used to examine the factor structure of the BPFSC. The hypothesized 4-factor structure was not supported. The unidimensional IRT analysis showed instances of local dependence among item pairs and item responses that were not strongly related to the underlying construct. As a consequence, items were eliminated, creating a unidimensional 11-item brief BPFSC (the BPFSC-11). Next, evidence of construct validity of scores based on the shortened version was evaluated using a different sample of 371 inpatient adolescents. We demonstrated similar indices of construct validity as observed for the BPFSC total score with the BPFSC-11 scores and found evidence for good criterion validity. Use of the BPFSC-11 in clinical settings will reduce the burden on respondents without loss of information.

Keywords: borderline personality features, adolescents, item response theory, factor analysis

Borderline personality disorder (BPD) is a serious psychological disorder characterized by a pattern of instability in interpersonal relationships, self-image, and affects, and marked impulsivity (American Psychiatric Association, 2013). Given the significant costs associated with adult BPD, there has been an increase in research examining borderline traits and disorder in adolescence to aid early identification and intervention (Sharp & Bleiberg, 2007; Skodol et al., 2002; Zanarini, Frankenburg, Hennen, & Silk, 2004). Good evidence now exists in support of the diagnosis of juvenile BPD (see Bondurant, Greenfield, & Tse, 2004, Chanen, 2012, Miller, Muehlenkamp, & Jacobson, 2008, and Sharp & Romero, 2007, for reviews), and the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*; American Psychiatric Association, 2013) allows for the diagnosis of BPD in adolescents.

Despite these advances, there remains a serious dearth of studies that examine the psychometric properties of measures to assess maladaptive personality functioning in youth (Sharp, Ha, Michonski, Venta, & Carbone, 2012). Early identification and treatment of BPD is dependent on the careful and accurate assessment of personality pathology in adolescents. Valid and reliable instruments that are both time and cost effective may helpfully complement clinical assessment. Such instruments should take into account the recent emphasis on dimensional models of maladaptive personality functioning (Helzer, Kraemer, & Krueger, 2006), especially when youth are concerned, as categorical approaches to personality assessment focus narrowly on clinically relevant symptoms, and do not allow for the study of the entire range of borderline symptoms (Crick et al., 2005). Assessment of symptoms across the full latent trait of borderline personality allows identification of not only those individuals who are demonstrating clinically significant levels of symptomatology but also those who may be considered at risk. Therefore, consistent with the developmental psychopathology principles of homotypic and heterotypic continuity, dimensional approaches allow for the characterization of all possible developmental trajectories toward or away from psychopathology over time as children mature through adolescence into early adulthood.

The BPFSC is the only dimensional measure, to date, specifically developed to assess borderline personality features in chil-

dren and adolescents. Crick et al. (2005) developed this self-report instrument by modifying the borderline (BOR) scale of the Personality Assessment Inventory (PAI; Morey, 1991), which is a reliable and valid tool used to assess borderline personality features among adults. Although an adolescent version of the PAI was developed, its items remained largely unchanged from the adult version. The BPFSC included age-appropriate items to reflect the original four domains of the PAI (affective instability, identity problems, negative relationships, and self-harm). In a community sample of 400 fourth through sixth graders, Crick et al. established evidence for the construct validity and moderate stability of the BPFSC. Evidence for criterion validity against an interview-based diagnostic measure of BPD in inpatient adolescents (Chang, Sharp, & Ha, 2011), and concurrent validity in a community sample of boys (Sharp, Mosko, Chang, & Ha, 2011), have also been established.

As yet, the performance of the individual BPFSC items using current full-information item response theory (IRT) approaches has not been examined in any age group drawn from either the community or clinical settings. When using IRT to examine item performance, it is necessary to first establish unidimensionality of a scale. Against this background, our first aim was to examine the putative four-factor structure of the BPFSC (affective instability, identity problems, negative relationships, self-harm/impulsivity) through a confirmatory multidimensional IRT approach, and we expected valuable information to be gleaned about the individual item functioning of the 24 BPFSC item set. Although this was necessary in order to evaluate individual item performance, conducting a factor analysis was also important for two additional reasons. First, from a practical and clinical standpoint, evidence for the four factors would imply that subscale scores can be used to identify subgroups of adolescents high on one or more dimensions but lower on others. Second, from a conceptual point of view, questions have been raised about the underlying factor structure of BPD, with several studies now suggesting a unidimensional factor structure for BPD as opposed to the more traditionally accepted three- or four-factor structure (Michonski, Sharp, Steinberg, & Zanarini, 2013). Although the Cronbach's alpha for the BPFSC has fallen into an "acceptable" range (.76 in the original validation sample [Crick et al., 2005] and .83 in a recent clinical sample [Sharp et al., 2012]), a high Cronbach's alpha is not indicative of a single unidimensional latent construct (e.g., Cortina, 1993). Therefore, more research is needed in evaluating the factor structure of the BPFSC, specifically, and the borderline construct in adolescence, more generally.

Past research (e.g., Gibbons et al., 2008) suggests that IRT item analysis of lengthy measures results in a reduction of the number of items via a selection of items with optimal characteristics; reliability may be even increased using fewer items (e.g., Steinberg & Thissen, 1996). This is advantageous for the use of the BPFSC in large-scale epidemiological and developmental studies of maladaptive personality functioning and for reducing the burden on respondents in clinical assessment settings. In anticipation of the shortening of the BPFSC, our second aim was to evaluate the construct validity of the new, shorter BPFSC by examining its performance against the original 24-item BPFSC version in a clinical sample of adolescent inpatients. First, we examined the capacity of a shortened BPFSC to distinguish between adolescents with and without BPD, as defined in the fourth edition of the *DSM*

(*DSM-IV-TR;* American Psychiatric Association, 2000), as determined by semistructured clinical interview through ROC analyses. Next, we examined correlations between scores obtained from the long and shorter versions of the BPFSC, and scores from other measured constructs known to relate to BPD in adults and adolescents, that is, emotion dysregulation and self-harm 1; Gratz & Roemer, 2004).

Method

Participants

Community sample for factor analysis and assessment of individual item performance. Nine hundred sixty-four adolescents between the ages of 14 and 19 years who participated in the second wave of an ongoing longitudinal study ($n = 1,042$ at Wave 1) were included in the present study. Participants were originally recruited from seven public schools representing five major school districts in a large and diverse metropolitan city in the United States. Study recruitment occurred during school hours in classes with mandated attendance. All students present in the selected classes were eligible to participate. Parental permission forms, in both English and Spanish, were sent home with students for their parents to review, sign, and return for a $5 gift card regardless of whether or not they were granted parental permission to participate. Of the 1,702 students present on recruitment days, 1,215 returned parental permission forms (71%), 1,119 obtained parental permission to participate (66% of those recruited; 92% of those who returned permission forms), and 1,046 completed the survey (62% of those recruited; 94% of those who received parental permission). Four surveys were discarded due to overt random responding, which resulted in 1,042 participants recruited and assessed. Participants were 55.9% female ($n = 539$) and 44.1% were male ($n = 425$). The age breakdown for the sample was as follows: 0.4% were 14 years old, 23% were 15, 50% were 16, 23% were 17, and 3.5% were 18 and older. Racial identification was as follows: 31.7% Hispanic, 30.3% White, 26.6% African American, 3.4% Asian/Pacific Islander, and 8% Other.

Clinical sample for construct validity. The sample included consecutive admissions to the adolescent treatment program of a private tertiary-care inpatient treatment facility specializing in the evaluation and stabilization of patients who failed to respond to previous interventions. Of the 454 adolescents admitted, 83 were excluded for a variety of reasons, including a desire not to participate in research ($n = 32$), discharge before assessments were completed ($n = 12$), active psychosis, IQ <70, diagnosis of autism spectrum disorder ($n = 37$), and primary language not being English. Thus, the final sample size consisted of 371 adolescents, of whom 33% ($n = 123$) met criteria for BPD as determined by a semistructured-interview-based diagnostic measure (see Measures section). Of the full sample, 52.6% suffered from a mood disorder, 7.7% from eating disorders, 43.4% from externalizing disorders and 54.7% from anxiety disorders. The ethnic breakdown of the sample was as follows: 91.8% White, 6.4% Hispanic, 4.5% Asian, 1.4% bi- or multiracial, and 2.3% Black, and the socioeconomic status of the sample was primarily upper middle class.

Measures

BPFSC. The BPFSC (Crick et al., 2005) was developed from the BPD scale (BOR) of the PAI (Morey, 1991), and consists of 24 items measuring borderline personality features in childhood (for ages 9 and older, including adolescents). These items assess how participants feel about themselves and other people, and are rated on a 5-point Likert-type scale ranging from *not true at all* to *always true*. Items comprise four subscales reflective of core BPD features: Affective Instability, Identity Problems, Negative Relationships, and Self-Harm. Sample items include "I want to let some people know how much they've hurt me" and "When I'm mad, I can't control what I do." Construct validity has been established, and borderline personality features detected by the BPFSC have been shown to be moderately stable across time (Crick et al., 2005).

Childhood Interview for *DSM-IV-TR* Borderline Personality Disorder (CI-BPD). The CI-BPD was used to examine the construct validity of the new shortened BPFSC. The CI-BPD (Zanarini, 2003) is a semistructured interview developed specifically for use with children and adolescents to assess BPD. The interview was adapted from an adult assessment of *DSM-IV-TR* personality disorders, with items modified from the borderline module of the Diagnostic Interview for *DSM-IV* Personality Disorders (DIPD-IV; Zanarini, Frankenburg, Sickel, & Yong, 1996). A total of nine criteria reflecting the *DSM-IV-TR* symptoms of BPD are rated using "0" for symptoms that are *absent*, "1" if the symptom is *probably present*, or "2" for symptoms that are *definitely present*. A minimum of five criteria scored as a 2 is required for a full diagnosis of BPD. Training on the CI-BPD involved several stages, which occurred in the following order: (a) didactic training by the first author (who was trained by the developers of the CI-BPD), (b) shadowing of interviews, (c) a practice interview with a nonpatient, (d) an interview with experienced interviewer critiquing, and, finally, (e) independent interviews. All interviews were video recorded. Monthly consensus meetings were held as booster sessions, in which video recordings were reviewed as a group to assure fidelity to the interview-based measure. Excellent psychometric properties have been reported for this measure (Sharp et al., 2012; Zanarini, 2003). Interrater reliability was calculated for 12% of the sample through obtaining ratings from two independent raters ($\kappa = .85$).

Difficulties in Emotion Regulation Scale (DERS). The DERS (Gratz & Roemer, 2004) is a 36-item self-report assessment of emotion dysregulation. Adolescents are asked to rate the frequency of each statement using a Likert-type scale ranging from $1 = $ *almost never* to $5 = $ *almost always*. A total score of emotion dysregulation is derived by summing all responses, covering several emotion regulation domains. These domains include (a) awareness and understanding of emotions; (b) acceptance of emotions; (c) the ability to engage in goal-directed behavior, and refrain from impulsive behavior, when experiencing negative emotions; (d) access to emotion regulation strategies perceived as effective; and (e) the flexible use of situationally appropriate strategies to modulate emotional responses. Higher scores indicate greater emotion regulation difficulties. Adequate psychometric properties of the measure have been reported (Gratz & Roemer, 2004; Perez, Venta, Garnaat, & Sharp, 2012). Reliability for this measure in the current sample was good (Cronbach's alpha = .95).

Deliberate Self-Harm Inventory (DSHI). The DSHI (Gratz, 2001) is a self-report measure containing 17 items assessing for the presence and frequency of various self-harm behaviors. Adequate psychometric properties have been reported, with high internal consistency and good test–retest reliability. In this study, we use the continuous total score on the DSHI to examine construct validity of the new, shortened BPFSC. High scores reflect a high frequency of self-harm and low scores indicate low frequency. Reliability for this measure in the current sample was good (Cronbach's alpha = .83).

Procedures

This study was approved by the appropriate institutional review board, and the data are part of a larger data set investigating adolescent health behaviors. Recruitment occurred during school hours in classes with required attendance. Research staff attended each class twice prior to assessment to explain the study and to answer questions. Information about the study, as well as parental permission slips, were sent home with the students for their parents to read, sign, and return. Assent was then obtained from students who returned the forms, and those who assented were pulled from class into a private room on campus to complete the survey.

Results

Measurement Models

The methods of IRT were used to evaluate the factor structure of the 24-item BPFSC. The IRT model fitting and the computation of the test statistics were performed using IRTPRO 2.1(Cai, du Toit, & Thissen, 2011). Goodness of fit of the IRT 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), and the standardized local dependence (LD) chi-square statistics (based on the LD statistic proposed by Chen & Thissen, 1997). The graded response model (Samejima, 1969, 1997) was selected as the item response model for these analyses; the graded model has often been found useful for questionnaire data collected using Likert-type scales (for examples, see Fraley, Waller, and Brennan (2000), Gray-Little, Williams, and Hancock (1997), and Steinberg, 1994, 2001).

Evaluating the Factor Structure of the 24-Item BPFSC

Multidimensional model. Our first analysis is designed to mirror the conceptual factors of the PAI-BOR as described in Crick et al. (2005) using the methods of full-information bifactor analysis. Item bifactor analysis refers to a type of confirmatory multidimensional IRT model in which one general factor and one or more specific factors are specified (Cai, Yang, & Hansen, 2011). Table 1 presents the IRT slope parameters and standard errors for the full-information item bifactor model. In this analysis, one general factor (all 24 items) and four specific factors (affective instability, identity problems, negative relationships, and self-harm/impulsiveness) are modeled. In IRT, the slope parameters are algebraically related to factor loadings and represent the degree of relation to the underlying construct. Slope parameters that are 1 or

Table 1
Slope Parameter Estimates and Standard Errors for the Bifactor and Unidimensional IRT Models

| Item | Abbreviated content | Bifactor model | | | | | | | | | | Unidimensional model | |
		a_1	SE	a_2	SE	a_3	SE	a_4	SE	a_5	SE	a	SE
8	feelings are strong	**2.38**	0.20	**3.24**	0.10	0.00	—	0.00	—	0.00	—	**1.27**	0.08
15	get into trouble . . . do w/o thinking	**2.37**	0.40	0.00	—	0.00	—	0.00	—	**2.45**	0.48	**1.62**	0.10
14	back and forth between feelings	**2.00**	0.12	0.21	0.09	0.00	—	0.00	—	0.00	—	**1.90**	0.11
19	upset, do things . . . aren't good	**1.98**	0.13	0.00	—	0.00	—	0.00	—	0.33	0.09	**1.95**	0.13
17	when mad, can't control what I do	**1.96**	0.13	0.55	0.11	0.00	—	0.00	—	0.00	—	**1.95**	0.13
21	get so mad, I can't let all anger out	**1.82**	0.12	0.40	0.10	0.00	—	0.00	—	0.00	—	**1.76**	0.11
18	how I feel about myself changes	**1.82**	0.11	0.00	—	0.33	0.09	0.00	—	0.00	—	**1.70**	0.11
13	People . . . let me down	**1.69**	0.15	0.00	—	0.00	—	**0.96**	0.28	0.00	—	**1.52**	0.10
9	something important missing	**1.65**	0.10	0.00	—	0.10	0.09	0.00	—	0.00	—	**1.60**	0.10
7	do things without thinking	**1.63**	0.15	0.00	—	0.00	—	0.00	—	**2.09**	0.29	**1.27**	0.09
16	worry that people will leave . . .	**1.56**	0.10	0.00	—	0.58	0.10	0.00	—	0.00	—	**1.49**	0.10
10	friends . . . treated me badly	**1.54**	0.13	0.00	—	0.00	—	0.84	0.30	0.00	—	**1.43**	0.10
11	careless with things	**1.43**	0.11	0.00	—	0.00	—	0.00	—	0.40	0.09	**1.45**	0.10
20	friends and I are mean to each other	**1.32**	0.12	0.00	—	0.00	—	−0.16	0.21	0.00	—	**1.28**	0.09
2	feel very lonely	**1.31**	0.09	0.00	—	0.00	—	0.19	0.15	0.00	—	**1.25**	0.09
6	let people know . . . hurt me	**1.15**	0.12	0.00	—	0.00	—	**0.90**	0.35	0.00	—	**1.06**	0.08
3	upset when . . . parents/friends leave	**1.01**	0.13	0.00	—	**2.78**	0.17	0.00	—	0.00	—	0.61	0.07
4	do things . . . wild/out of control	**0.96**	0.08	0.00	—	0.00	—	0.00	—	0.74	0.09	**1.00**	0.08
12	change mind about what to do . . .	**0.90**	0.08	0.00	—	0.25	0.08	0.00	—	0.00	—	**0.90**	0.08
1 (r)	pretty happy person	0.83	0.08	−0.34	0.08	0.00	—	0.00	—	0.00	—	0.74	0.07
22 (r)	I get bored very easily	0.78	0.07	0.00	—	−0.04	0.08	0.00	—	0.00	—	0.81	0.07
24 (r)	once a friend, we stay friends	0.36	0.07	0.00	—	0.00	—	0.02	0.19	0.00	—	0.34	0.06
23 (r)	take good care of [my] things	0.27	0.07	0.00	—	0.00	—	0.00	—	0.26	0.08	0.27	0.06
5 (r)	feel . . . same way all the time	0.08	0.06	−0.06	0.08	0.00	—	0.00	—	0.00	—	0.05	0.06

Note. Boldface values indicate slope parameter values of .90 or higher. Dashes indicate that there are no standard errors for fixed parameters. IRT = item response theory; a = slope parameters; b = threshold parameters; *SE* = standard errors.

greater are typically considered substantial; however, in Table 1, we adopted a more conservative approach that includes slope values of .90 or higher (indicated in bold). The column labeled *a1* is the general factor that includes all 24 items. The columns labeled *a2*, *a3*, *a4*, and *a5* include the items for the specific factors, namely, affective instability, identity problems, negative relationships, and self-harm/impulsiveness, respectively. (The column labeled "a" lists the slope parameters for a separate analysis that will be described later.)

A perusal of the column for the general factor (*a1*) shows that the majority of the items had substantial slope parameter estimates; however, the five reverse-scored items had low slope parameter estimates, indicating little relation to the underlying construct defined by the items included in the analysis. The specific factors (*a2* through *a5*) reveal two different kinds of patterns of slope parameters. The first pattern represents what has been termed a "doublet" in the psychological measurement literature and can be described as LD. LD occurs when items are more strongly correlated than can be accounted for by the general factor. The specific factors *a4* (negative relationships) and *a5* (self-harm/impulsiveness) show LD. For example, the two items comprising the content of self-harm/impulsiveness, Item 7 ("I do things without thinking") and Item 15 ("I get into trouble because I do things without thinking") are similar in wording and meaning. The LD is most likely a result of common wording. These two items are the only ones to show substantial slope parameters on that specific factor. Thus, this is not the "self-harm/impulsiveness" construct intended with the six-item set, but rather excess covariation that cannot be accounted for by the general factor. Similarly, the specific factor

"negative relationships" (*a4*) is defined by LD between the Item 6 ("I want to let some people know how much they've hurt me") and Item 13 ("People who are close to me have let me down"). The second pattern of slope parameters associated with the specific factors represents unique item variance. The specific factors *a2* (identity problems) and *a3* (negative relationships) each have only a single substantial slope parameter; this indicates unique variance that is not accounted for by the general factor.

The bifactor model also showed 12 pairs of items with substantial LD (values of 10 or greater are considered noteworthy). The standardized LD chi-square statistics imply that the bifactor model is not adequate to account for the excess covariation between these item pairs.

The full-information bifactor analysis indicated that the pattern of item covariation is multidimensional. However, the patterns of slope parameters for the specific factors exhibit LD and unique variance, rather than being reflective of individual differences on the intended constructs.

Unidimensional models. As a consequence of the bifactor model showing some items that did not have a substantial relation to the underlying construct, and that many of the items showed LD, we now focused on developing a shorter unidimensional version of the BPFSC. The next IRT analysis evaluates the magnitude of the item slope parameters by specifying a unidimensional model. Our aim is to select items that show substantial slope parameters for inclusion in the short version. The two rightmost columns of Table 1 list the slope parameters and their associated standard errors for the 24 items. Again, about a quarter of the items do not show a substantial relation to the underlying construct, and

there are 16 pairs of items that show LD; the standardized chi-square LD values ranged from 10.0 to 51.3. The substantial LD values indicate that a unidimensional model is not adequate to account for item covariation. Next, we evaluated a unidimensional model that includes 18 of the BPFSC items that showed substantial slope parameters in the 24-item unidimensional model analysis.

The graded model item parameters are shown in Table 2. The slope parameters, representing the degree of relation of the item responses to the underlying construct as defined by the 18 BPFSC items, shows substantial slopes for all items except Item 12 ("I change my mind almost everyday about what I should do when I grow up"). However, seven item pairs exhibit LD (the standardized chi-square values ranged from 10.1 to 38.4). In most cases, the item with the lower slope parameter was considered for omission. In other cases, with such a small item set, we focused on item content in addition to item parameters to guide item selection. On the basis of LD with one or more items, Items 7, 10, and 17 were omitted. We eliminated three additional items for the following reasons: (a) Item 12 has a low slope parameter; (b) Item 4 ("I do things that other people would consider wild or out of control") brings in the concept of the perceptions of others, and it is double barreled in the sense that "wild" is not the same as "out of control"; and (c) the phrase "can't let my anger out" in Item 21 ("I get so mad I can't let my anger out") is ambiguous (e.g., the phrase could have any of the following meanings: such anger might be uncontrollable if the respondent allowed its expression; it would be inappropriate to express such an amount of anger; the respondent is not able to express anger).

An intermediate IRT analysis was completed to evaluate the performance of the remaining 12 BPFSC items. The results of this analysis revealed two item pairs exhibiting LD, specifically, Item Pairs 18 ("How I feel about myself changes a lot ") and 14 ("I go back and forth between different feelings, like being mad or sad or happy"), and 19 ("When I get upset, I do things that aren't good for me") and 15 ("I get into trouble because I do things without thinking"). On the basis of similarity of content, Item 19 was omitted for the next analysis. We retained Items 14 and 18 because the content reflects different aspects of borderline personality features; because the presence of LD depends on the item set, we will reevaluate these two items in the context of the 11-item analysis.

Table 3 presents the slope and threshold parameters for the 11 retained items. The graded IRT model showed satisfactory fit (M_2 [869] = 2151.64, $p < .001$; RMSEA = 0.04); however, Items 14 and 18 show LD (standardized χ^2 LD index = 10.6). An analysis was done to evaluate the significance of the LD. Specifically, a bifactor model that includes an equal-slope second factor composed of the item pair (Items 14 and 18) that showed LD was used. The estimated slope parameter for the specific factor is 0.69, with a standard error of 0.18. To evaluate the significance of the LD, a likelihood ratio goodness-of-fit difference test was done, subtracting −2 log likelihood obtained from the bifactor analysis from the −2 log likelihood from the eight-item unidimensional analysis. The result is G^2 (1) = 6.98, $p = .008$.

In practice, possible ways to deal with significant LD include (a) omitting one of the items in the pair, as we have done previously; or (b) retaining both items and forming a testlet (Steinberg & Thissen, 1996; Thissen & Steinberg, 2010) of the item pair by summing the item responses, thereby creating a single "super" item. The testlet is then used for item parameter estimation (so that the slope parameters are not influenced by the excess covariation between the two items showing LD).

We retained the 11 items for the short version of the BPFSC by using the testlet comprised of the sum of the item responses for Items 14 and 18. Table 4 lists the item parameters for the now 10-item analysis (nine items, one testlet). Because the testlet is made from the sum of the two items, total scores can be calculated by summing up the responses to the 11 items. Thus, the testlet accounts for the local dependence in the item analysis without altering the calculation of the summed score.

The test information curve is shown in the upper panel of Figure 1. Test information curves show how well the construct is mea-

Table 2

Slope and Threshold Parameter Estimates and Standard Errors for 18 BPFSC Items

Abbreviated content	a	SE	b_1	SE	b_2	SE	b_3	SE	b_4	SE
2. feel very lonely	1.20	0.08	−0.96	0.09	0.50	0.07	1.74	0.12	3.05	0.21
4. do things . . . wild	1.01	0.08	−0.93	0.10	0.45	0.08	2.06	0.16	3.79	0.31
6. let people know	1.07	0.08	−1.62	0.13	−0.42	0.08	1.00	0.09	2.15	0.16
7. do things w/o	1.26	0.09	−1.75	0.13	−0.41	0.07	1.33	0.09	2.61	0.17
8. feelings are strong	1.30	0.09	−1.66	0.12	−0.50	0.07	0.76	0.07	1.73	0.11
9. something . . . missing	1.59	0.10	−1.12	0.09	−0.25	0.06	0.76	0.06	1.49	0.09
10. friends . . . badly	1.43	0.10	−0.24	0.06	0.86	0.07	2.06	0.13	2.96	0.20
11. careless with things	1.44	0.10	−0.18	0.06	0.97	0.07	2.05	0.13	2.98	0.20
12. change mind	0.89	0.08	−0.39	0.09	1.25	0.12	2.78	0.24	3.97	0.35
13. people . . . let me down	1.54	0.10	−1.28	0.09	−0.27	0.06	0.97	0.07	1.81	0.11
14. back and forth	1.91	0.12	−0.80	0.07	0.17	0.05	1.05	0.06	1.76	0.09
15. trouble . . . do w/o think	1.63	0.10	−0.64	0.07	0.23	0.05	1.40	0.08	2.19	0.13
16. worry . . . people . . . leave	1.48	0.10	−0.76	0.08	0.09	0.06	0.97	0.07	1.75	0.11
17. mad, can't control	1.99	0.13	0.02	0.05	0.90	0.06	1.68	0.09	2.22	0.12
18. feel . . . myself changes	1.69	0.11	−0.70	0.07	0.31	0.05	1.41	0.08	2.25	0.13
19. upset, do things	2.00	0.13	−0.07	0.05	0.82	0.05	1.68	0.09	2.32	0.13
20. friends and I are mean	1.28	0.10	−0.13	0.07	1.29	0.09	2.63	0.18	3.67	0.29
21. can't let all anger out	1.77	0.12	−0.22	0.06	0.60	0.05	1.36	0.08	2.09	0.12

Note. BPFSC = Borderline Personality Features Scale for Children; a = slope parameters; b = threshold parameters; SE = standard errors.

Table 3

Slope and Threshold Parameter Estimates and Standard Errors for 11 BPFSC Items

Item content	a	SE	b_1	SE	b_2	SE	b_3	SE	b_4	SE
2. I feel very lonely.	1.36	0.09	−0.89	0.08	0.45	0.06	1.61	0.11	2.81	0.19
6. I want to let some people know how much they've hurt me.	1.21	0.09	−1.50	0.12	−0.40	0.07	0.90	0.08	1.98	0.14
8. My feelings are very strong. For instance, when I get mad, I get really really mad. When I get happy, I get really really happy.	1.19	0.08	−1.77	0.13	−0.55	0.08	0.78	0.08	1.83	0.13
9. I feel that there is something important missing about me, but I don't know what it is.	1.86	0.12	−1.05	0.07	−0.25	0.05	0.69	0.06	1.38	0.08
11. I'm careless with things that are important to me.	1.20	0.09	−0.21	0.07	1.07	0.09	2.31	0.16	3.39	0.26
13. People who were close to me have let me down.	1.67	0.11	−1.23	0.09	−0.27	0.06	0.92	0.07	1.73	0.10
14. I go back and forth between different feelings, like being mad or sad or happy.	2.26	0.14	−0.76	0.06	0.14	0.05	0.96	0.06	1.65	0.08
15. I get into trouble because I do things without thinking.	1.20	0.09	−0.79	0.09	0.25	0.07	1.66	0.12	2.65	0.19
16. I worry that people I care about will leave and not come back.	1.69	0.11	−0.72	0.07	0.06	0.05	0.89	0.07	1.62	0.09
18. How I feel about myself changes a lot.	1.81	0.12	−0.69	0.07	0.28	0.05	1.35	0.08	2.18	0.12
20. Lots of times, my friends and I are really mean to each other.	1.01	0.09	−0.17	0.08	1.50	0.13	3.12	0.25	4.39	0.39

Note. BPFSC = Borderline Personality Features Scale for Children; *a* = slope parameters; *b* = threshold parameters; *SE* = standard errors.

sured at all levels of the underlying construct continuum. Measurement precision is approximately constant for values of the construct continuum between −1 and +2. Total information for the 11 BPFSC items is approximately 8 for this range of the continuum. The standard errors of IRT scores for this range are approximately $\frac{1}{\sqrt{8}}$ = 0.35; this translates to an IRT approximation of reliability of .875 (calculated as one minus the measurement variance) for the BPFSC-11 scores in that range. For comparison, the traditional reliability estimates using Cronbach's alpha for the 24-item and the 11-item BPFSC scores are .88 and .85, respectively.

Using the methods of IRT, we have developed an 11-item version of the BPFSC. In the next section, we investigate evidence of construct validity for the short version compared with the long version.

Construct Validity

We examined construct validity of scores based on the shortened version (BPFSC-11) compared with the original 24-item BPFSC. The first analysis investigates BPFSC group mean differences between adolescents that met criteria for BPD compared with those who did not meet the criteria for BPD on the CI-BPD. Results of independent sample *t* tests demonstrated significant group mean differences, $t(369) = -10.57$, $p < .001$, for the

24-item BPFSC: Adolescents with BPD had significantly higher scores ($M = 80.56$, $SD = 14.21$) compared with nonborderline psychiatric controls ($M = 64.23$, $SD = 14.21$). These differences were replicated with the BPFSC-11, $t(369) = -10.23$, $p < .001$. The BPD group had significantly higher scores ($M = 38.34$, $SD = 6.91$) compared with nonborderline psychiatric controls ($M = 29.88$, $SD = 38.34$). Thus, scores obtained with the BPFSC-11 were found to show the expected group differences in borderline traits previously demonstrated with scores based on the entire 24-item BPFSC.

We used receiver operating characteristic (ROC) analyses to further determine the clinical utility of the BPFSC-11, including its sensitivity and specificity. ROC curves for the BPFSC-11 and CI-BPD diagnosis are shown in Figure 2.

ROC analyses showed that area under the curve (AUC) values and standard errors were significant ($p < .05$). The BPFSC-11 had an AUC of .80, indicating good diagnostic accuracy. Sensitivity and specificity analyses demonstrated an ideal cutoff point of 34 (sensitivity = .740; specificity .= 714) on the BPFSC-11. We also examined the diagnostic accuracy of the 24-item BPFSC, which demonstrated an AUC value of .80 as well.

Next, correlational analyses were conducted with the DERS (emotion dysregulation) and DSHI (self-harm) total scores. Results

Table 4

Graded Model Item Parameter Estimates for the BPFSC-11 Items, Using a Testlet Combining Items 14 and 18

Item	a	SE	b_1	SE	b_2	SE	b_3	SE	b_4	SE	b_5	SE	b_6	SE	b_7	SE	b_8	SE
2	1.36	0.09	−0.89	0.08	0.45	0.06	1.61	0.11	2.81	0.19								
6	1.22	0.09	−1.49	0.11	−0.40	0.07	0.90	0.08	1.97	0.13								
8	1.19	0.09	−1.77	0.13	−0.55	0.08	0.78	0.08	1.83	0.13								
9	1.87	0.12	−1.05	0.07	−0.25	0.05	0.69	0.06	1.38	0.08								
11	1.22	0.09	−0.21	0.07	1.06	0.09	2.29	0.16	3.35	0.25								
13	1.68	0.11	−1.23	0.09	−0.28	0.06	0.91	0.07	1.73	0.10								
15	1.20	0.09	−0.79	0.09	0.25	0.07	1.66	0.12	2.65	0.19								
16	1.71	0.11	−0.72	0.07	0.06	0.05	0.88	0.06	1.62	0.09								
20	1.03	0.09	−0.17	0.08	1.49	0.13	3.08	0.25	4.34	0.39								
Testlet 14 and 18	2.48	0.15	−1.14	0.07	−0.66	0.05	0.00	0.05	0.43	0.05	0.99	0.06	1.39	0.07	1.89	0.09	2.26	0.11

Note. BPFSC-11 = 11-Item Borderline Personality Features Scale for Children; *a* = slope parameters; *b* = threshold parameters; *SE* = standard errors.
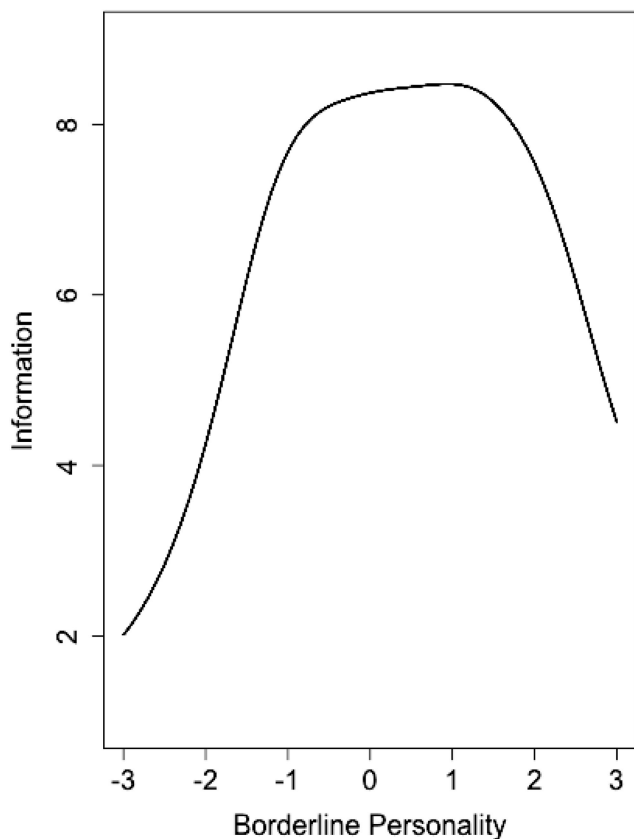
Figure 1. Test information curve for the BPFSC-11 showing how well the construct is measured at all levels of the underlying construct continuum.

showed that the pattern of correlations between measures of emotion dysregulation and self-harm and the BPFSC ($r = .71$, $p < .001$, and $r = .67$, $p < .001$, respectively) is replicated with the BPFSC-11 ($r = .67$, $p < .001$, and $r = .42$, $p < .001$, respectively).

Discussion

A need has been identified for the development and evaluation of measures to assess personality pathology in adolescents. Against this background, the current study aimed to examine the factor structure and individual item performance of the BPFSC using current full-information approaches in a large community sample of adolescents. In so doing, we refined the 24-item BPFSC and evaluated the concurrent and criterion validity of the BPFSC-11 in an inpatient sample of adolescents. Results demonstrated expected correlations with emotion dysregulation and self-harm, and the BPFSC-11 showed good sensitivity and specificity in identifying adolescents who meet *DSM-IV-TR* criteria for BPD. In all, we provide evidence in support of a short and easy-to-administer unidimensional measure of BPD that shows good construct validity.

The full-information bifactor analysis indicated that the pattern of item covariation of the BPFSC item set is multidimensional. However, close inspection of the pattern of slope parameters

suggested that the BPFSC exhibits a type of multidimensionality that is more reflective of local dependence and unique variance than the measurement of individual differences on intended psychological constructs. Thus, although the aim of the current article was not to compare and contrast different factor models underlying the BPD construct, the findings from the bifactor analysis did not support the original putative four-factor structure of the BPFSC. Instead, when items with low slope and LD were removed to form the BPFSC-11, support for a unidimensional factor structure for the BPFSC-11 was found. This is consistent with several studies that have demonstrated a unidimensional factor structure for BPD assessment tools in adults (e.g. Aggen, Neale, Roysamb, Reichborn-Kjennerud, & Kendler, 2009; Feske, Kirisci, Tarter, & Pilkonis, 2007; Sanislow et al., 2002), children (for example, Michonski et al., 2013), and adolescents (for example, Sharp et al., 2012). Evidence of dimensionality depends on the pattern of covariation among the items; theoretically, it is possible that multiple factors might be found with items specifically written to measure indicators of the more narrowly defined constructs.

Our findings have two important practical implications for the use of the BPFSC both clinically and in research. First, the lack of support of a four-factor structure for the BPFSC recommends against the use of the four subscale scores. Second, although construct validity was demonstrated solely for inpatient settings in the current study, the BPFSC-11, given its brevity, holds promise for use in large-scale epidemiological studies as well as resource-constrained clinical settings, pending further research on the psychometric properties of the BPFSC-11 in other settings. In clinical settings, the BPFSC-11 can be used as brief screening tool to justify further, more in-depth clinical assessment.

The assessment of personality pathology in adolescence is important, given its prevalence in both inpatient and outpatient clin-
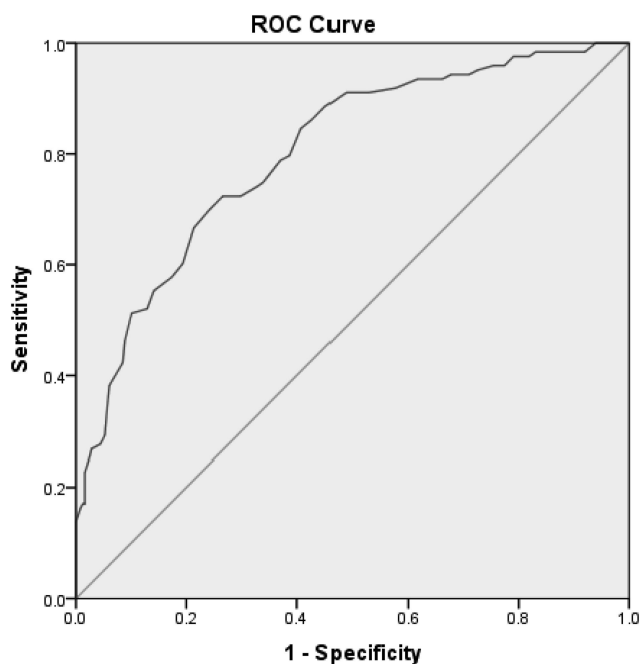

Figure 2. ROC curve for BPFSC-11 in detecting BPD as defined by the CI-BPD.

ical settings. BPD is commonly misdiagnosed or missed completely, particularly in the adolescent population, in which emotional dysregulation and externalizing behavior can be explained as developmentally appropriate, depending on its magnitude (Sharp & Bleiberg, 2007). When evaluating patients, both mental health and primary care practitioners who make the diagnosis of BPD have relied more heavily on external, self-destructive behaviors such as nonsuicidal self-injury, recurrent suicide attempts, and impulsive acts, as pathognomonic characteristics of the disorder are lacking (Treloar & Lewis, 2009). The availability and routine use of a brief screen for BPD could potentially permit for earlier detection of BPD in clinical samples, perhaps prior to the development of more severe externalizing indicators. The availability of a brief and reliable screen for BPD would increase the likelihood that practitioners alerted to the possibility of BPD would pursue a more formal evaluation (Noblin, Venta, & Sharp, 2013; Patel, Sharp, & Fonagy, 2011).

The current study has several limitations. As mentioned, results for the BPFSC-11 generalize only to inpatient settings, and future research must replicate these findings in community and other clinical settings. The item content of the BPFSC in general, and in the BPFSC-11, is somewhat limited, in that it includes no explicit coverage of self-harm. The adult and adolescent versions of the PAI-BOR on which the BPFSC was originally modeled includes only one item that vaguely refers to self-harm ("When I'm upset, I typically do something to hurt myself") and may therefore be similarly limited. Notwithstanding these limitations, we hope that the introduction of a brief and easy-to-administer screening measure of BPD would stimulate research and build on the current momentum in the study of adolescent personality pathology.

References

Aggen, S. H., Neale, M. C., Roysamb, 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, 39,* 1967–1978. doi:10.1017/S0033291709005807

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bondurant, H., Greenfield, B., & Tse, S. M. (2004). Construct validity of the adolescent borderline personality disorder: A review. *The Canadian Child and Adolescent Psychiatry Review, 13,* 53–57.

Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Chicago, IL: Scientific Software 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

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods, 16,* 221–248. doi:10.1037/a0023350

Chanen, A. (2012). Review: Urgent need for RCT evidence on effectiveness of crisis interventions for borderline personality disorder [Comment]. *Evidence Based Mental Health, 15,* 94. doi:10.1136/eb-2012-100928

Chang, B., Sharp, C., & Ha, C. (2011). The criterion validity of the Borderline Personality Feature Scale for Children in an adolescent inpatient setting. *Journal of Personality Disorders, 25,* 492–503. doi:10.1521/pedi.2011.25.4.492

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.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104.

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

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

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78,* 350–365. doi:10.1037/0022-3514.78.2.350

Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagiolini, A., Grochocinski, V. J., . . . Immekus, J. C. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatric Services, 59,* 361–368. doi:10.1176/appi.ps.59.4.361

Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the deliberate self-harm inventory. *Journal of Psychopathology and Behavioral Assessment, 23,* 253–263. doi:10.1023/A:1012779403943

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 26,* 41–54. doi:10.1023/B:JOBA.0000007455.08539.94

Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An item response theory analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 23,* 443–451. doi:10.1177/0146167297235001

Helzer, J. E., Kraemer, H. C., & Krueger, R. F. (2006). The feasibility and need for dimensional psychiatric diagnoses. *Psychological Medicine, 36,* 1671–1680. doi:10.1017/S003329170600821X

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

Michonski, J. D., Sharp, C., Steinberg, L., & Zanarini, M. C. (2013). 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. *Personality Disorders: Theory Research, and Treatment, 4,* 15–22. doi:10.1037/a0027948

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

Morey, L. (1991). *Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources.

Noblin, J. L., Venta, A., & Sharp, C. (2013). The validity of the MSI-BPD among inpatient adolescents. *Assessment*. Advance online publication. doi:10.1177/1073191112473177

Patel, A. B., Sharp, C., & Fonagy, P. (2011). Criterion validity of the MSI-BPD in a community sample of women. *Journal of Psychopathology and Behavioral Assessment, 33,* 403–408. doi:10.1007/s10862-011-9238-5

Perez, J., Venta, A., Garnaat, S., & Sharp, C. (2012). The difficulties in Emotion Regulation Scale: Factor structure and association with non-

suicidal self-injury in adolescent inpatients. *Journal of Psychopathology and Behavioral Assessment, 34,* 393–404. doi:10.1007/s10862-012-9292-7

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society. Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of item response theory* (pp. 85–100). New York, NY: Springer-Verlag.

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. *The 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 & F. Volkmar (Eds.), *Lewis' child and adolescent psychiatry: Comprehensive textbook* (pp. 680–691). Baltimore, MD: Lippincott Williams and Wilkins.

Sharp, C., Ha, C., Michonski, J., Venta, A., & Carbone, C. (2012). Borderline personality disorder in adolescents: Evidence in support of the Childhood Interview for *DSM-IV* Borderline Personality Disorder in a sample of adolescent inpatients. *Comprehensive Psychiatry, 53,* 765–774. doi:10.1016/j.comppsych.2011.12.003

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.

Skodol, A. E., Gunderson, J. G., McGlashan, T. H., Dyck, I. R., Stout, R. L., Bender, D. S., . . . Oldham, J. M. (2002). Functional impairment in patients with schizotypal, borderline, avoidant, or obsessive-compulsive personality disorder. *The American Journal of Psychiatry, 159,* 276–283. doi:10.1176/appi.ajp.159.2.276

Steinberg, L. (1994). Context and serial-order effects in personality measurement: Limits on the generality of measuring changes the measure. *Journal of Personality and Social Psychology, 66,* 341–349. doi:10.1037/0022-3514.66.2.341

Steinberg, L. (2001). The consequences of pairing questions: Context effects in personality measurement. *Journal of Personality and Social Psychology, 81,* 332–342. doi:10.1037/0022-3514.81.2.332

Steinberg, L., & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods, 1,* 81–97. doi:10.1037/1082-989X.1.1.81

Thissen, D., & Steinberg, L. (2010). Using item response theory to disentangle constructs at different levels of generality. In S. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 123–144). Washington, DC: American Psychological Association. doi:10.1037/12074-006

Treloar, A. J. C., & Lewis, A. J. (2009). Diagnosing borderline personality disorder: Examination of how clinical indicators are used by professionals in the health setting. *Clinical Psychologist, 13,* 21–27. doi:10.1080/13284200802392536

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

Zanarini, M. C., Frankenburg, F. R. S., Sickel, A. E., & Yong, L. (1996). *The Diagnostic Interview for DSM-IV Personality Disorders.* Belmont, MA: McLean Hospital, Laboratory for the Study of Adult Development.

Zanarini, M. C., Frankenburg, F. R., Hennen, J., & Silk, K. R. (2004). Mental health service utilization by borderline personality disorder patients and Axis II comparison subjects followed prospectively for 6 years. *Journal of Clinical Psychiatry, 65,* 28–36. doi:10.4088/JCP.v65n0105