Assessing Psychopathy Among Justice Involved Adolescents with the PCL: YV: An Item Response Theory Examination Across Gender

Siny Tsang,
University of Virginia

Karen M. Schmidt,
University of Virginia

Gina M. Vincent,
University of Massachusetts Medical School

Randall T. Salekin,
University of Alabama

Marlene M. Moretti, and
Simon Fraser University

Candice L. Odgers
Duke University

Abstract

This study used an item response theory (IRT) model and a large adolescent sample of justice involved youth (N = 1,007, 38% female) to examine the item functioning of the Psychopathy Checklist – Youth Version (PCL: YV). Items that were most discriminating (or most sensitive to changes) of the latent trait (thought to be psychopathy) among adolescents included "Glibness/ superficial charm", "Lack of remorse", and "Need for stimulation", whereas items that were least discriminating included "Pathological lying", "Failure to accept responsibility", and "Lacks goals." The items "Impulsivity" and "Irresponsibility" were the most likely to be rated high among adolescents, whereas "Parasitic lifestyle", and "Glibness/superficial charm" were the most likely to be rated low. Evidence of differential item functioning (DIF) on four of the 13 items was found between boys and girls. "Failure to accept responsibility" and "Impulsivity" were endorsed more frequently to describe adolescent girls than boys at similar levels of the latent trait, and vice versa for "Grandiose sense of self-worth" and "Lacks goals." The DIF findings suggest that four PCL: YV items function differently between boys and girls.

Adolescents in the juvenile justice system routinely undergo psychological testing and assessment using a vast array of instruments, one of which is the Psychopathy Checklist: Youth Version (PCL: YV; Forth, Kosson, & Hare, 2003). The PCL: YV represents a downward extension of a widely used assessment tool among adults, the Psychopathy Checklist (PCL and PCL-R; Hare, 1980; 1991; 2003). Psychopathy is characterized by a number of interpersonal, affective, and behavioral features, such as lack of remorse and guilt, callousness, and superficial charm (Cleckley, 1976; Hare, 1991).

Many of the items from the PCL: YV were transferred down from the PCL-R with little alteration, leading some to question the appropriateness of the use of this scale with adolescents (Edens, Skeem, Cruise, & Cauffman, 2001). Prior research using the PCL: YV has provided some promising estimates of the reliability and validity of the instrument (see Hare, 2003 and Salekin & Lynam, 2010), however, findings related to the predictive validity of the PCL: YV have been considerably mixed. For instance, while many studies have found that the PCL: YV is predictive of antisocial conduct among adolescents (see Leistico, Salekin, DeCoster, & Rogers, 2008; Olver, Stockdale, & Wormith, 2009 for a review), other research has shown that PCL: YV scores do not predict future offending, especially among adolescent girls (e.g., Odgers, Reppucci, & Moretti, 2005; Vincent, Odgers, McCormick, & Corrado, 2008). The inabilty of the PCL-YV to predict future offending among adolescents, and girls specfically, has caused some observers to question the usefulness of the instrument with adolescents in the juvenile jutice system in general, and, more specifically, has raised questions about whether the PCL: YV may be measuring something different in boys versus girls (Odgers, Moretti, & Reppucci, 2005). If the measure is to continue to be used clinically and with juvenile justice involved youth, more information will be needed on the item functioning of this and other psychopathy indices.

In this study we examine the item functioning of the PCL: YV among one of the largest samples of adolescents offenders assembled to date (N =1,007) using Item Response Theory (IRT) methods. The application of IRT offers the advantages of identifying specific items that may be especially useful in capturing the underlying construct of psychopathy among this population, testing whether items may function differentially, or demonstrate bias, across different populations. IRT also facilitates the identification of items that are good at distinguishing among individuals with different levels of the latent trait, which in this case is thought to be psychopathy.

The factor structure of the Psychopathy Checklist (PCL; Hare, 1980) and the PCL-R (Hare, 1991; 2003) has been studied extensively (e.g., Cooke & Michie, 2001; Hare, 2003; Harpur, Hare, and Hakstian, 1989; Vitacco, Rogers, Neumann, Harrison, & Vincent, 2005; Vitale, Smith, Brinkley, & Newman, 2002). Initially described as having a two-factor (Hare, 1980; 1991), a four-factor model (Bolt, Hare, Vitale, & Newman, 2004; Hare, 2003) and a 13-item, three-factor model (Cooke & Michie, 2001) have also been proposed. Similar controversies regarding the most appropriate factor structure apply to the PCL: YV (e.g.,

Jones, Cauffman, Miller, & Mulvey, 2006; Kosson et al., 2013; Neumann, Kosson, Forth, & Hare, 2006; Salekin, Brannen, Zalot, Leistico, & Neumann, 2006; Sevecke, Pukrop, Kosson, & Krischer, 2009).

Item Response Theory

Certain features of IRT are particularly advantageous in examining the functioning of individual items of the PCL: YV. The S-shaped form of Item Characteristic Curves (ICCs) describes the probability of item responses regressed on trait levels. The location and slope of the ICCs describe the varying item difficulty and discrimination, respectively, across items. The location of an ICC corresponds to the latent trait level at which the probability of having a "correct" or an "incorrect" item response is equivalent. Item difficulty (β) reflects the extent to which the probabilities of item responses differ across trait levels. A smaller β value implies a higher probability of a "correct" item response at lower trait levels, indicating an easier to endorse (i.e., observed relatively more frequently) item. The ICC slope (a) reflects the extent to which item response probabilities vary as a function of differences in the latent trait. The steeper the slope, the more likely item response probabilities change with small variation in latent trait. An item with a larger a value is better at differentiating across changes in trait levels, indicating a more discriminating item (Crocker & Algina, 2008; Embretson & Reise, 2000).

Item properties of the PCL: YV

To date, only three studies have examined the item properties of the PCL: YV using IRT (Dillard, Salekin, Barker, & Grimes, 2012; Schrum & Salekin, 2006; Vincent, 2002). Using Samejima's (1969) graded response model (GRM) with 13 items of the PCL: YV, Vincent (2002) tested for the presence of age-related measurement bias between adolescent male offenders and a large sample of adult male offenders. Results showed many items, particularly the behavioral items, were less discriminating at various levels of the latent trait for adolescents than for adults, suggesting age-related measurement bias primarily among the socially deviant behavioral items.

Schrum and Salekin (2006) were the first to utilize the GRM to examine the test and item functioning of all the 20 PCL: YV among a sample of adolescent girls ($N = 123$). Results showed that "callous/lack of empathy", "conning/manipulative", and "grandiose sense of self-worth" were the most discriminating items, whereas "poor anger control", "shallow affect", and "serious violations of conditional release" were the least discriminating items among the detained adolescent girls. There existed considerable variation of item difficulty among the PCL: YV items; "impulsivity", "poor anger control", and "juvenile delinquency" were the most frequently endorsed by clinicians rating adolescents, implying that the probability of receiving a high score on these items was high even for girls with low levels of the latent trait.

More recently, Dillard et al. (2012) explored the differential item functioning (DIF) of the 20 PCL: YV items among a sample of male ($n = 307$) and female ($n = 144$) juvenile offenders. Consistent with Schrum and Salekin (2006), "conning/manipulative" was found to be one of the most discriminating items for both boys and girls. However, there was DIF

between boys versus girls for 11 of the 20 items. For example, "shallow affect" was found to be more discriminating for boys than for girls, whereas "impersonal sexual behavior" and "lacks goals" were better at differentiating among girls with varying levels of the latent trait than boys. These findings suggest that the PCL: YV items may demonstrate gender bias when assessing the latent psychopathy trait between adolescent boys and girls, which should be taken into consideration in the interpretation of the PCL: YV scoring. Nonetheless, further replication of these findings is needed to further our understanding of the item functioning of the PCL: YV across gender.

To the best of our knowledge, all published IRT studies on the PCL: YV and its adult counterpart, the PCL-R, applied Samejima's (1969) graded response model (GRM) (e.g., Bolt et al., 2004; Cooke & Michie, 1997, 1999; Cooke, Kosson, & Michie, 2001; Dillard et al., 2012; Hare, 2003; Schrum & Salekin, 2006). Although the GRM is applicable for items with ordered categorical responses, the GRM assumes that the probabilities of being rated in a higher category response increase as the latent psychopathy trait level increases. For a particular item, an individual with low latent trait levels should be less likely to receive a high category response (e.g., 1 vs. 0) than one with high latent trait levels. However, this *hypothesized* order may not be reflected in empirical data (Andrich, 2012). If the likelihoods of response categories do not increase with the latent trait level, at least one response option is never the most probable response, conditional on trait level (Andrich, 1988). For an item with "disordered thresholds", the probability of receiving a higher category response (e.g., 1 vs. 0) may actually be *higher* for the individual with low latent trait levels. The GRM's assumption of ordered category responses implies that the GRM is unable to detect "disordered threshold" cases.

In order to verify the order of category responses empirically, the generalized partial credit model (G-PCM; Muraki, 1993) can be used. The G-PCM, while assuming ordered category responses, does not force the values of the thresholds to be ordered, hence allowing the detection of "disordered thresholds." The G-PCM is also appropriate for analyzing personality scale responses in which raters assess the extent to which items are useful in characterizing and scoring levels of a latent trait among individuals and, by extension, testing whether the items and the assessment instrument may demonstrate bias when applied within different populations. Further details are described in the methods section.

The Current Study

The purpose of the present study was two-fold. We aimed to examine how well the PCL: YV items functioned within a large sample of adolescent offenders (N =1,007), and to test for potential DIF between justice involved boys and girls. This latter aim is important as much of the prior research has been limited in the ability to test for differential item functioning between boys and girls due to the limited sample sizes of girls (Dillard et al., 2012).

Method

Participants

Combining three research samples, this study utilized data from a total of 1,007 adolescent offenders. Sample 1 consisted of 327 post-adjudication youth, recruited from maximum and minimum custody settings from British Columbia (Vincent, 2002). Sample 2 ($n = 241$) consisted of 120 youth from British Columbia (Penney & Moretti, 2007), and 121 from Virginia (Odgers, Moretti & Reppucci, 2010). At the time of data collection, 60 were in an assessment unit, and 181 were post adjudication. Sample 3 ($n = 439$) included 139 adolescents from Florida and 300 from Alabama. At the time of data collection, 139 were at a court evaluation unit, and 300 were at pretrial detention (Dillard et al., 2012).[1] The combined sample consisted of 615 boys and 392 girls, with ages ranging from 12 to 18 years. The majority of participants self-identified as Caucasian, with smaller percentages of Black, Native American/Aboriginal Canadian, and Hispanic. As shown in Table 1, there were no substantial differences between genders.

Measures

The latent trait of psychopathy was assessed with the Psychopathy Checklist: Youth Version (PCL: YV; Forth et al., 2003) across all three studies. The PCL: YV consists of 20 items, designed for use with adolescents ranging from 12 to 18 years old. The PCL: YV was scored by trained researchers, based on a semi-structured interview with the youth and a review of available collateral information (e.g., school reports, criminal history). Each item was scored on a 3-point ordinal scale (0 = *does not apply*; 1 = *applies to a certain extent*; 2 = *applies*). A total PCL: YV score can be obtained by summing the 20 items, ranging from 0 to 40.

Generalized Partial Credit Model (G-PCM; Muraki, 1993)—Assuming that item i is scored $x = 0, \ldots, m_i$ for an item with $K_i = m_i + 1$ response categories, the category response curves (CRCs), indicating the probability of a person to be scored in category x on a m_i step item conditional on trait level, are computed as:

$$P_{ix}(\theta) = \frac{exp\sum_{j=0}^{x}\alpha_i(\theta - \delta_{ij})}{\sum_{r=0}^{M}[exp\sum_{j=0}^{r}\alpha_i(\theta - \delta_{ij})]}$$

where $\sum_{j=0}^{0}\alpha_i(\theta - \delta_{ij}) = 0$; α_i indicates the slope parameter for item i; δ_{ij} ($j = 1, \ldots, m_i$) indicates the intersection point of two CRCs. At low latent trait levels, a person is most likely to get a score of 0 (*does not apply*). As the latent trait level increases along the scale continuum, the probability of receiving a score of 0 decreases, whereas the probability of receiving a score of 1 (*applies to a certain extent*) increases. At high levels of the latent trait, a score of 2 (*applies*) is the most likely outcome. The category intersection points (δ_{i1}, δ_{i2}) indicates the corresponding trait levels where it is equally likely to get either score (i.e., 0 or 1, 1 or 2).

---

[1]Item functioning of the PCL: YV in sample 3 has previously been published (Dillard et al., 2012). The findings were compared and discussed in the Discussion section.

The slope parameters a_i indicate the extent of response category variations across items with changes in the latent trait level. Items with larger a_i values change more rapidly in item response probabilities across different trait levels (Embretson & Reise, 2000). The G-PCM will allow us to identify relatively effective items for measuring the latent trait of psychopathy, which may provide information regarding which characteristics are more important indicators of psychopathy within this sample. All item parameters in the G-PCM were estimated on a log-odds scale, using maximum likelihood in PARSCALE 4.1 (Muraki & Bock, 2003).

Differential item functioning—Concerns have been raised about whether the PCL: YV can be effectively used to assess psychopathy among adolescent boys versus girls (Dillard et al., 2012; Schrum & Salekin, 2006). If the relations between items and the latent trait of psychopathy are not equivalent between adolescent boys and girls, the measurement of the latent trait may not be "invariant" between gender groups. These items indicate differential item functioning (DIF), meaning the probabilities of being scored in different response categories are not the same for boys versus girls, despite having the same underlying trait level. Trait level estimates may be biased if the same item parameters are used for both gender groups (Embretson & Reise, 2000); as a result, a boy and a girl with the same underlying trait level of psychopathy may receive different scores on some PCL: YV items. Thus, clinicians and researchers may be generating biased PCL: YV scores that may, depending on the extent of the bias, render subsequent comparisons with respect to levels of psychopathy and subtypes across gender invalid.

To examine DIF, a multi-group IRT parameter calibration was conducted. Constraining the slope (a_i) parameters to be the same for boys and girls, the category intersection (δ_{ij}) parameters were estimated simultaneously, with boys serving as the reference group. The focal group adjustment approach (e.g., see Embretson & Reise, 2000) used in this study does not require anchoring items; all items in both groups can be assessed simultaneously. After adjusting for trait level differences (theta), significant differences in the category intersection parameters (δ_{ij}) indicate DIF on items. This approach is different from the anchoring method (Thissen, Steinberg, & Gerrard, 1986) used in prior studies (e.g., Cooke & Michie, 1999), which focuses on some k items while other ($k-1$) items serve as anchor items.

Results

Descriptive Statistics

Item descriptive statistics for the entire sample, and by gender, are presented in Table 2. As there remains some debate about the factor structure for the PCL: YV, with studies showing similar fit indices for the three-factor and four-factor models (see Jones et al., 2006; Neumann et al., 2006; Salekin et al., 2006; Vincent, 2002), we examined the model fit across models that imposed different factor structures using confirmatory factor analyses. Results showed that the 13-item, 3-factor version of the PCL: YV represented the best fit to the data (χ^2 (62) = 338.21, CFI = .89, TLI = .87, RMSEA = .09), thus the subsequent main analyses were conducted on the 13 PCL: YV items.[2] The three factors are hypothesized to

assess Arrogant and deceitful personality style (F1), Deficient affective experience (F2), and Impulsive and irresponsible lifestyle (F3) (Cooke & Michie, 2001).[3]

Unidimensionality is an assumption of IRT analyses, indicating that there exists one underlying latent construct influencing item responses within the measure (Embretson & Reise, 2000). To examine whether the 13-item PCL: YV version should be treated as a unidimensional structure or three separate factors, the dimensionality of the PCL: YV was examined using confirmatory factor analysis (CFA) and multi-dimensional IRT (MIRT) (in package 'mirt'; Chalmers, 2013). The 3-factor model (CFA results: CFI = .96, TLI = .97, RMSEA = .06, χ^2 = 240.28; MIRT results: log likelihood = −11793.61, AIC = 23671.21, BIC = 23878.29) consistently showed a better fit than the 1-factor model (CFA results: CFI = .94, TLI = .93, RMSEA = .08, χ^2 = 481.94; MIRT results: log likelihood = −11924.88, AIC = 23927.76, BIC = 24120.05). Subsequent analyses were conducted separately for each of the three factors.

G-PCM estimates

As previously described, the G-PCM estimates the slope parameters (a) as well as the category intersection parameters (δ_{ij}) of the PCL: YV items. The slope parameter estimates indicate the extent to which categorical responses vary as a function of differences in the latent trait. The intersection points (δ_{i1}, δ_{i2}) reflect the corresponding latent psychopathy level at which it is equally likely to get either score (0 or 1, 1 or 2).[4]

The slope parameter estimates (a) indicated that the extent to which items are sensitive to varying latent trait levels of psychopathy were not uniform across the items within factors (Table 3). The a's were, on average, larger for F1: Arrogant & deceitful personality style (a's range from .59 to 1.18) and F2: Deficient affective experience items (a's range from .46 to 1.59). Among them, "Glibness/superficial charm" (a = 1.18) and "Lack of remorse" (a = 1.59) were the most sensitive to changes across different latent psychopathy traits. The CRCs of these items were relatively steep, reflecting that small variations in the latent trait level were more likely to result in changes of the response categories (Figure 1). On the other hand, F3: Impulsive and irresponsible lifestyle items, with smaller slope parameters (a's range from .24 to .87), were less sensitive to changes across varying latent psychopathy traits. In particular, the corresponding CRCs of "Lacks goals" (a = .24) and "Parasitic

---

[2] Confirmatory factor analyses were conducted on the (a) 20-item, 1 factor model (χ^2 (170) = 1346.79, CFI = .82, TLI = .80, RMSEA = .09), (b) Hare's (2003) 18-item, 4 factor model, where two of the items do not load onto any of the 4 factors (χ^2 (129) = 717.82, CFI = .91, TLI = .90, RMSEA = .07), (c) Cooke and Michie's (2001) 13-item, 3 factor model (χ^2 (62) = 338.21, CFI = .94, TLI = .92, RMSEA = .07), and (d) 13-item, 1 factor model, assuming unidimensionality for the 13 items (χ^2 (65) = 641.41, CFI = .87, TLI = .85, RMSEA = .10).

[3] In the absence of evidence supporting predictive validity (e.g., information about future antisocial behavior, arrest records), the examination of whether the proposed factor structure(s) is valid for the PCL: YV is beyond the scope of the current study. The CFAs provided here aim to identify the best factor model for subsequent IRT analyses.

[4] The GRM was also applied to examine the item properties of the PCL: YV items in the current study. Model fitness was examined using the AIC and BIC. Compared to the G-PCM (AICs = 7411.43, 7641.68, 9566.42; BICs = 7461.39, 7694.64, 9637.35, for F1, F2, and F3, respectively), results showed that sometimes the G-PCM fits better, and sometimes the GRM fits better for the data (AICs = 7408.402, 7640.45, 9566.64; BICs = 7458.36, 7693.40, 9637.58, and for F1, F2, and F3, respectively). The small differences in fit indices between the G-PCM and GRM suggests the GRM is not necessarily a better fit for the data (Kass & Raftery, 1995). Due to limited space, results of the GRM were not reported; however, all estimated parameters were similar, and the DIF results in the same direction, with nearly identical if not identical magnitudes between the GRM and G-PCM.

lifestyle" ($a = .42$) were relatively flat; indicating larger variation in the latent trait was needed for changes in the response categories.

The category intersection parameters (δ_{ij}) indicated that F3 items, with lower values, were observed relatively more frequently. More individuals were rated as 2 or 1, instead of 0 for these items. However, F1 were found to be observed relatively less frequently, with higher average values of δ_{ij}'s. More participants received lower (a score of 0) rather than higher (a score of 2) ratings, indicating individuals needed to have higher values on the latent trait in order to receive higher ratings of these items. The δ_{ij}'s of the items within factors were relatively well spread across the latent trait, meaning that increase in the latent trait level was reflected with the increased likelihood of receiving higher ratings on the items.

Differential Item Functioning Across Boys versus Girls

To examine whether the estimated latent trait levels were comparable for justice involved boys versus girls, the 13 PCL: YV items were examined for DIF, with boys as the reference group and girls as the comparison group. We fixed the slope parameter (a) to be the same for boys and girls and allowed the category intersection parameters (δ_{ij}) to be free to vary. Results showed some variability in item functioning between boys and girls in the PCL: YV items (Table 4). The CRCs of selected items, plotted separately for boys and girls, were shown in Figure 2.

For items that showed similar item functioning between genders, the changes in the likelihoods in response categories, and the probabilities of being rated in a particular response category, were similar for adolescent boys and girls. The CRCs of boys and girls for these items overlapped almost perfectly, demonstrating almost invariant item functioning between genders.

Significant differences in item functioning were found for "Grandiose sense of self-worth", "Failure to accept responsibility", "Lack goals", and "Impulsivity". With boys as the reference group, the CRCs of "Failure to accept responsibility" and "Impulsivity" were shifted slightly to the right for boys (Figure 2, bottom panel). For these items, when boys and girls were at similar latent trait levels, it was more likely for girls to receive a high rating (2 vs. 1, and 1 vs. 0) than boys. These average item scores, in raw scale units, were slightly higher for girls than boys (Table 2). Characteristics of "Failure to accept responsibility" and "Impulsivity" were rated as "applicable" more frequently for girls than boys. With increased probability of receiving higher scores even at lower levels of the latent trait, these features were observed relatively *more frequently* for girls than for boys. Conversely, the CRCs of "Grandiose sense of self-worth" and "Lacks goals" were shifted to the right for girls (Figure 2, bottom panel). At a given latent trait level, boys were more likely to be rated higher on these items than girls. These average item scores, in raw scale units, were slightly higher for boys than girls (Table 2). "Sense of grandiosity" and "Lack goals" were more likely to be rated as "applicable" for boys than girls; these characteristics were observed relatively *more frequently* for boys than girls.

Results from the DIF analyses suggested that, for a given latent trait level, the likelihood of getting rated in a particular response category (0, 1, or 2) may not be the same between boys

and girls for four of the 13 PCL-YV items. Girls were more likely to be perceived to be impulsive and fail to accept responsibility than boys, whereas boys were more likely to be rated as lacking goals and being grandiose.

Discussion

The purpose of this study was to evaluate the item properties and functioning of the PCL: YV for justice involved boys and girls within one of the largest samples of adolescents assembled to date. Results showed that there were slight differences in discrimination abilities across the PCL: YV items. Items assessing arrogant and deceitful personality style and deficient affective experience were relatively better at discriminating different levels of the psychopathy trait, meaning that adolescents with different psychopathy trait levels are likely to receive different scores (higher scores for those with stronger endorsement of psychopathy, and vice versa) on these items. On the other hand, impulsive and irresponsible lifestyle items were relatively less discriminating, suggesting that similar scores may be given to adolescents, regardless of their level of psychopathy trait. Findings also indicated that impulsive and irresponsible lifestyle items were endorsed by clinicians to describe adolescents relatively more frequently, suggesting that even adolescents low on the psychopathy trait had a relatively high probability of being rated as impulsive and irresponsible. For arrogant and deceitful personality style (F1) items, higher ratings (i.e., these features observed relatively less frequently) required having a higher latent psychopathy trait level. However, impulsive and irresponsible characteristics may be more frequently endorsed characteristics among adolescents, and more so among the juvenile offenders within these samples. Therefore, it may be reasonable that impulsive and irresponsible behavior are more common (e.g., Cauffman, Steinberg, & Piquero, 2005; Monahan, Steinberg, Cauffman, & Mulvey, 2009; Vitacco, Neumann, Robertson, & Durrant, 2002) due to the developmental stage of adolescence or the characteristics of adolescents for whom the PCL: YV is typically used (e.g., juvenile offenders). Due to the relative homogeneity of impulsive and irresponsible characteristics, slight changes in the latent trait may not be as readily reflected by changes in the response categories. On the other hand, parasitic lifestyle may be relatively uncommon among adolescents in general, and thus was less frequently identified among youth in this study.

The DIF analysis indicated gender-related variability in category intersection parameters, with four of the thirteen items showing significant DIF. Some items ("Failure to accept responsibility" and "Impulsivity") were more commonly used by raters to describe girls versus boys at similar trait levels, whereas other items ("Grandiose sense of self-worth" and "Lack goals") were more commonly endorsed for boys versus girls with similar trait levels. Among adolescents with similar latent trait levels, raters characterized girls as more irresponsible and impulsive, whereas boys were more likely to be assessed as having features of grandiose sense of self-worth and lacking goals (again versus girls at similar trait levels). Compared to boys with similar latent trait levels of psychopathy, it is possible that raters observed certain features indicative of the latent psychopathy trait relatively less frequently in girls. Although such biases may reflect differential gender norms for boys and girls among raters, it seems unlikely that all raters across the three samples were consistently

biased in their scorings on these specific behaviors. However, further research is warranted to examine potential rater effects on the PCL: YV.

It is also recommended that results from the current DIF analysis be interpreted with caution. Some PCL: YV items showed weak communality and weak factor loadings (e.g., "Lacks goals," "Failure to accept responsibility for own actions"), suggesting weak coherence within the hypothesized underlying factors (Table 3). We note that extracting meaningful comparisons of DIF for these items under such conditions may not be viable. The gender-related variability found should be interpreted as preliminary findings and for guiding refinements to the instrument. Results reported here require replication in other samples to render such gender differences reliable.

Limitations

Despite the large sample size and novel application of IRT methods to examine the functioning of the PCL: YV, this study also had several limitations. First, the extent to which the PCL: YV measures the latent trait of psychopathy among adolescents remains unclear. As with all psychological assessment instruments, there exists discrepancy between the latent construct and the operationalization of measurement; one cannot not be absolutely certain how well the PCL: YV items actually measure the "real", underlying construct of psychopathic personality. Second, the current study did not take into account potential variations in predictive outcomes for different items. The inclusion of outcome variables (e.g., offending) would have allowed us to evaluate the possibility that items differed in their prediction of various outcome variables. Such findings could potentially help identify the "best" or "most crucial" items for assessing psychopathy among adolescents. Third, differences in the base rates of items may have reflected differential gender norms for adolescent boys and girls. As the current study merged samples from diverse locations, it is possible that the training of interviewers and/or the reliability of coding might have differed across sites (although training on the assessment of the PCL: YV was conducted by the same individual for 3 of the 4 sites). However, the assessments of rater bias and/or potential site effects were beyond the scope of the present study, such effects should be examined in future studies.

Implications

Gender related differential functioning of PCL: YV items raised questions as to whether the total scores are indicative of the similar underlying latent trait levels for girls and boys, and more importantly, whether the PCL: YV, was measuring psychopathy among justice involved adolescent boys versus girls in the same way. It is possible that the opposing direction of DIF across items may cancel out when computing the total score, such that the overall effect of DIF on the total PCL: YV scores could be negligible (Cooke et al., 2001). One may argue that such a view fails to consider that items are indicative of different PCL: YV factors and thus gender related bias may result in inaccurate interpretation of factor scores or the test profile. This point is well taken; however, within the existing assessment framework, the effect of item functioning differences on the expected total PCL: YV scores are expected to be only modest. Nonetheless, given the importance placed on assessing psychopathy in adolescents, particularly in juvenile justice contexts, further research is

required to better understand the similarity and difference in the construct of psychopathy between genders.

An important issue for future research will be to examine whether certain items are better (or worse) in predicting particular outcomes (e.g., violence, recidivism) between adolescent boys and girls. If the construct of psychopathy manifests in different behavior characteristics among boys and girls, it is possible that items that are more discriminating for one gender may also be more predictive of negative outcomes for that group.

Results from the current study also highlighted the need to improve the assessment of the latent trait of psychopathy using the PCL: YV. As many of the PCL: YV items were "inherited" from the PCL-R, the extent to which the features indicative of psychopathy in adults can be extended to adolescents remain unclear (see Vincent et al, 2008 for a fuller discussion of this point). First, we urge researchers to examine whether the PCL: YV items function differently between justice versus non-justice involved adolescents. Assuming the latent trait of psychopathy is much less prominent among adolescents outside of the justice system, items that are more sensitive to variations in the latent trait of psychopathy would be more informative in the assessment of psychopathy. Second, the current ordinal response categories of the PCL: YV may not be sensitive to changes in the features indicative of the psychopathy trait. It is also possible that the difference between categories 0 and 1 is much bigger (or smaller) than that between categories 1 and 2, and such differences may not be consistent across items and/or among raters, resulting in potential bias in the scoring of PCL: YV items. If the response categories can be revised to be more sensitive to changes in the observed characteristics (e.g., interval scaling), the reliability of the PCL: YV may be substantially improved.

Although previous studies utilized the GRM (e.g., Dillard et al., 2012; Schrum & Salekin, 2006) to examine item functioning of the PCL: YV items, their results were compared with the current findings using the G-PCM. Consistent with previous studies, "Glibness/ superficial charm or Impression management" and "Parasitic lifestyle" were also not commonly used by raters to describe offending adolescents in the current study (Dillard et al., 2012; Schrum & Salekin, 2006). We acknowledge these behaviors may reflect stable personality traits indicative of psychopathy among adults; however, as the personality traits of adolescents are still developing, it is possible that such characteristics are not sophisticated enough among delinquent youth to be apparent to raters, making them less likely to receive high scores on these items. Alternatively, these items may require more revision of the item descriptions in the manual in order to be developmentally appropriate. Consistent with previous findings, "Grandiose sense of self-worth" and "Callous/lack of empathy" were among the items that are most sensitively reflect changes in the latent psychopathy traits among offending girls than boys (Dillard et al., 2012; Schrum & Salekin, 2006). In addition, "Shallow affect" was found to be applied more frequently for justice involved boys versus girls, whereas "Failure to accept responsibility" was applied more frequently to girls versus boys at similar levels of the latent trait (Dillard et al., 2012). "Lacks goals" and "Impulsivity" were previously found to exhibit no substantial gender differences (Dillard et al., 2012), however, our findings with the G-PCM suggested that "Lacks goals" was more commonly applied to describe boys, whereas "Impulsivity" was

more commonly endorsed to describe girls versus boys. While the sources of the differences in findings are not known, our findings, consistent with prior studies, indicated that item properties of the PCL: YV items are not invariant. Further research is required to investigate DIF with a variety of IRT models in the PCL: YV among diverse samples of adolescents.

The current study is the first to utilize the G-PCM to examine the item properties of the PCL: YV. Although in the previous section, attempts were made to compare the present findings with existing literature, caution is warranted in such comparisons, as the two IRT models (G-PCM and GRM) are not directly comparable, particularly across samples. As such, future research is needed to examine whether or not item properties of the PCL: YV remain invariant across age groups. Longitudinal studies are also needed to explore whether or not item properties remain consistent between the PCL: YV and the adult PCL-R within samples. Although both instruments were developed to assess the construct of psychopathy, it is possible that certain items may be more relevant for identifying psychopathic features among adolescents, whereas others are better among adults.

References

Andrich D. A general form of Rasch's extended logistic model for partial credit scoring. Applied Measurement in Education. 1988; 1:363–378.10.1207/s15324818ame0104_7

Andrich D. An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any "threshold disorder controversy. Educational and Psychological Measurement. 2012; 73:78–124.10.1177/0013164412450877

Bolt DM, Hare RD, Vitale JE, Newman JP. A multigroup item response theory analysis of the Psychopathy Checklist--Revised. Psychological Assessment. 2004; 16:155–168.10.1037/1040-3590.16.2.155 [PubMed: 15222812]

Cauffman E, Steinberg L, Piquero A. Psychological, neuropsychological, and psychophysiological correlates of serious antisocial behavior in adolescence: The role of self-control. Criminology. 2005; 43:133–176.10.1111/j.0011-1348.2005.00005.x

Chalmers RP. mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software. 2012; 48:1–29.

Cleckley, H. The mask of sanity. 5. St. Louis, MO: Mosby; 1976.

Cooke DJ, Michie C. An item response theory analysis of the Hare Psychopathy Checklist--Revised. Psychological Assessment. 1997; 9:3–14.10.1037/1040-3590.9.1.3

Cooke DJ, Michie C. Psychopathy across cultures: North America and Scotland compared. Journal of Abnormal Psychology. 1999; 108:55–68.10.1037/0021-843X.108.1.58

Cooke DJ, Michie C. Refining the construct of psychopathy: Towards a hierarchical model. Psychological Assessment. 2001; 13:171–188.10.1037/1040-3590.13.2.171 [PubMed: 11433793]

Cooke DJ, Kosson DS, Michie C. Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist--Revised (PCL-R) in Caucasian and African American participants. Psychological Assessment. 2001; 13:531–542.10.1037/1040-3590.13.4.531 [PubMed: 11793896]

Crocker, L.; Algina, J. Introduction to classical and modern test theory. Mason, OH: Cengage Learning; 2008.

Dillard CL, Salekin RT, Barker ED, Grimes RD. Psychopathy in adolescent offenders: An item response theory study of the Antisocial Proces Screening Device-Self Report and the Psychopathy Checklist: Youth Version. Personality Disorders: Theory, Research, and Treatment. 2012

Edens JF, Skeem JL, Cruise KR, Cauffman E. Assessment of "juvenile psychopathy" and its association with violence: A critical review. Behavioral Sciences and the Law. 2001; 19(1):53–80.10.1002/bsl.425 [PubMed: 11241681]

Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah, NJ: Psychology Press; 2000.

Forth, AE.; Kosson, DS.; Hare, RD. The Psychopathy Checklist: Youth Version. Toronto: Multi-Health Systems; 2003.

Hare RD. A research scale for the assessment of psychopathy in criminal populations. Personality and Individual Differences. 1980; 1:111–119.10.1016/0191-8869(80)90028-8

Hare, RD. Manual for the Revised Psychopathy Checklist. 1. Toronto, Ontario, Canada: Multi-Health Systems; 1991.

Hare, RD. The Hare Psychopathy Checklist--Revised. 2. Toronto: Multi-Health Systems; 2003.

Harpur TJ, Hare RD, Hakstian AR. Two-factor conceptualization of psychopathy: Construct validity and assessment implications. Psychological Assessment: A Journal of Consulting and Clinical Psychology. 1989; 1:6–17.10.1037/1040-3590.1.1.6

Jones S, Cauffman E, Miller JD, Mulvey E. Investigating different factor structures of the Psychopathy Checklist: Youth Version: Confirmatory factor analytic findings. Psychological Assessment. 2006; 18:33–48.10.1037/1040-3590.18.1.33 [PubMed: 16594810]

Kosson DS, Neumann CS, Forth AE, Salekin RT, Hare RD, Krischer MK, Sevecke K. Factor structure of the Hare Psychopathy Checklist: Youth Version (PCL: YV) in adolescent females. Psychological Assessment. 2013; 25:73–83.10.1037/a0028986

Leistico AM, Salekin RT, DeCoster J, Roger R. A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. Law and Human Behavior. 2008; 32:28–45. [PubMed: 17629778]

Monahan KC, Steinberg L, Cauffman E, Mulvey EP. Trajectories of antisocial behavior and psychosocial maturity from adolescence to young adulthood. Developmental Psychology. 2009; 45:1654–1668.10.1037/a0015862 [PubMed: 19899922]

Muraki E. Information functions of the generalized partial credit model. Applied Psychological Measurement. 1993; 17:351–363.10.1177/014662169301700403

Muraki, E.; Bock, RD. PARSCALE 4 for Windows: IRT based test scoring and item analysis for graded and rating scales [Computer software]. Skokie, IL: Scientific Software International, Inc; 2003.

Neumann CS, Kosson DS, Forth AE, Hare RD. Factor structure of the Hare Psychopathy Checklist: Youth Version in incarcerated adolescents. Psychological Assessment. 2006; 18:142–154.10.1037/1040-3590.18.2.142 [PubMed: 16768590]

Odgers CL, Moretti MM, Reppucci ND. Examining the science and practice of violence risk assessment with female adolescents. Law and Human Behavior. 2005; 29(1):7–27.10.1007/s10979-005-1397-z [PubMed: 15865330]

Odgers CL, Moretti MM, Reppucci ND. A review of findings from the Gender and Aggression Project: Informing juvenile justice policy and practice through gender sensitive research. Court Review. 2010; 46:6–9.

Odgers CL, Reppucci ND, Moretti MM. Nipping psychopathy in the bud: An examination of the convergent, predictive, and theoretical utility of the PCL-YV among adolescent girls. Behavioral Sciences and the Law. 2005; 23:743–763.10.1002/bsl.664 [PubMed: 16333812]

Olver ME, Stockdale KC, Wormith JS. Risk assessment with young offenders: A meta-analysis of three assessment measures. Criminal Justice and Behavior. 2009; 36:329–353.

Penny SR, Moretti MM. The relation of psychopathy to concurrent aggression and antisocial behavior in high-risk adolescent girls and boys. Behavioral Sciences and the Law. 2007; 25:21–47.10.1002/bsl.715 [PubMed: 16953486]

Salekin RT, Brannen DN, Zalot AA, Leistico AM, Neumann CS. Factor structure of psychopathy in youth: Testing the applicability of the new four-factor model. Criminal Justice and Behavior. 2006; 33:135–157.10.1177/0093854805284416

Salekin, RT.; Lynam, DR. Child and adolescent psychopathy: An introduction. In: Salekin, RT.; Lynam, DR., editors. Handbook of Child and Adolescent Psychopathy. New York, NY: Guilford Press; 2010. p. 1-11.

Samejima, F. Psychometrika Monographs. Vol. 17. Richmond, VA: Psychometric Society; 1969. Estimation of latent ability using a response pattern of graded scores. Retrieved from http://www.psychometrika.org/journal/online/MN17.pdf

Schrum CL, Salekin RT. Psychopathy in adolescent female offenders: An item response theory analysis of the Psychopathy Checklist: Youth Version. Behavioral Sciences and the Law. 2006; 24:39–63.10.1002/bsl.679 [PubMed: 16491478]

Sevecke K, Pukrop R, Kosson DS, Krischer MK. Factor structure of the Hare Psychopathy Checklist-Youth Version in German felame and male detainees and community adolescents. Psychological Assessment. 2009; 21:45–56. [PubMed: 19290765]

Thissen D, Steinberg L, Gerrard M. Beyond group mean differences: The concept of item bias. Psychological Bulletin. 1986; 99:118–128.10.1037/00332909.99.1.118

Vincent, GM. Unpublished doctoral dissertation. British Columbia, Canada: Simon Fraser University; 2002. Investigating the legitimacy of adolescent psychopathy assessments: Contributions from item response theory.

Vincent GM, Odgers CL, McCormick AV, Corrado RR. The PCL: YV and recidivism in male and female juveniles: A follow-up into young adulthood. International Journal of Law and Psychiatry. 2008; 31:287–296. [PubMed: 18534679]

Vitacco MJ, Neumann CS, Robertson AA, Durrant SL. Contributions of impulsivity and callousness in the assessment of adjudicated male adolescents: A prospective study. Journal of Personality Assessment. 2002; 78:87–103.10.1207/S15327752JPA7801_06 [PubMed: 11936214]

Vitacco MJ, Rogers R, Neumann CS, Harrison KS, Vincent GM. A comparison of factor models on the PCL-R with mentally disordered offenders: The development of a four-factor model. Criminal Justice and Behavior. 2005; 32:526–545.10.1177/0093854805278414

Vitale JE, Smith SS, Brinkley CA, Newman JP. The reliability and validity of the Psychopathy Checklist--Revised in a sample of female offenders. Criminal Justice and Behavior. 2002; 29:202–231.10.1177/0093854802029002005
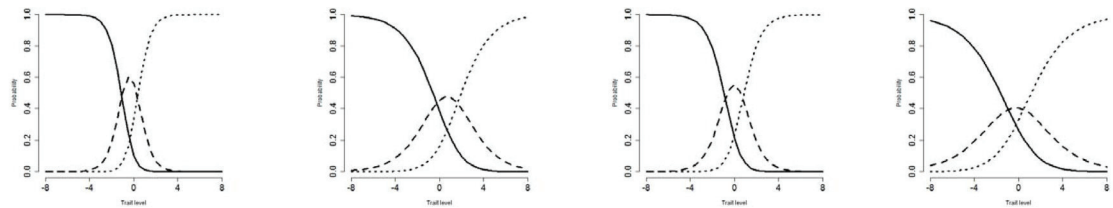
Factor 1

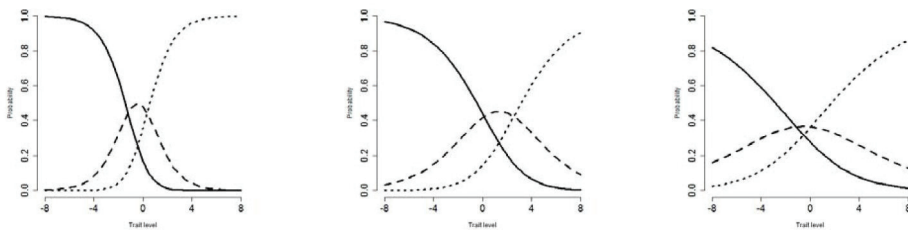1: Glibness/superficial charm   2: Grandiose sense of self-worth   4: Pathological lying   5: Conning/manipulative
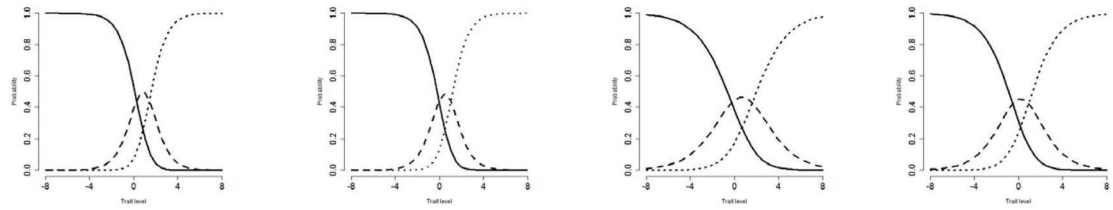
Factor 2

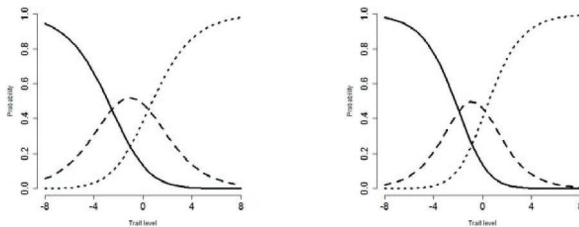6: Lack of remorse or guilt   7: Shallow affect   8: Callous/lack of empathy   16: Failure to accept responsibility

Factor 3

3: Need for stimulation   9: Parasitic lifestyle   13: Lacks goals
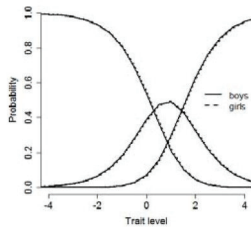
14: Impulsivity   15: Irresponsibility

Figure 1.
Category response curves of PCL: YV items. Solid line: 0 = does not apply; Dashed line: 1 = applies to a certain extent; Dotted line: 2 = applies.
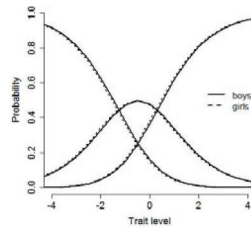Category response curves of PCL: YV 3 items. Solid line: 0 = does not apply; Dashed line: 1 = applies to a certain extent; Dotted line: 2 = applies.

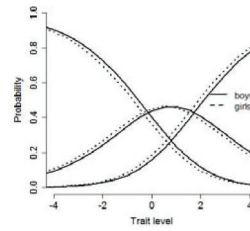Selected items with no differential item functioning
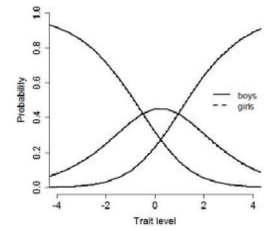
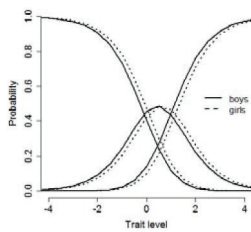| 1: Glibness/superficial charm | 3: Need for stimulation | 4: Pathological lying | 5: Conning/manipulative |

Items with differential item functioning

| 2: Grandiose sense of self-worth | 16: Failure to accept responsibility | 13: Lacks goals | 14: Impulsivity |
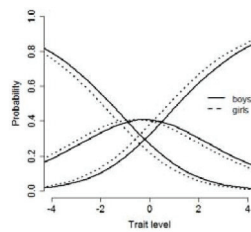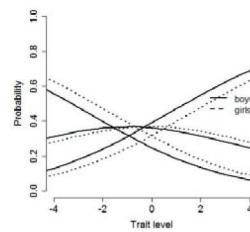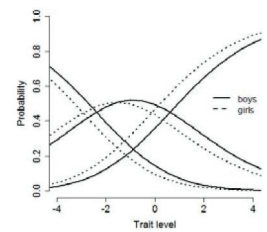
Figure 2.
Selected items with/without significant differential item functioning for boys (solid lines) and girls (dotted lines). Top panel: Selected items that showed no significant differential item functioning between boys and girls. Bottom Panel: Factor 2 and 3 items that showed significant differential item functioning between boys and girls. Items 2 and 13 were found to relatively more common for boys, as indicated by the slight shift of the CRCs to the left. Items 16 and 14 were found to be relatively less common for boys, as indicated by the slight shift of the CRCs to the right.

Table 1

Demographics of participants by gender

	Boys (*n* = 615)	Girls (*n* = 392)	Total (*N* = 1007)
Age	16.00 (*SD* = 1.59)	15.64 (*SD* = 1.47)	15.82 (*SD* = 1.61)
Ethnicity			
Caucasian	50.25%	39.27%	45.93%
Black	26.32%	38.74%	31.20%
Native Canadian/American	9.17%	11.26%	9.99%
Hispanic	8.32%	4.45%	6.80%
Other	5.94%	6.28%	6.07%

Table 2

Descriptive statistics of the Psychopathy Checklist: Youth Version by gender

	Boys [a]		Girls [b]		Total [c]	
	M	SD	M	SD	M	SD
1 Glibness/superficial charm	0.55	0.70	0.56	0.69	0.56	0.69
2 Grandiose sense of self-worth	0.75	0.78	0.68	0.71	0.73	0.76
3 Need for stimulation	1.22	0.69	1.19	0.77	1.21	0.72
4 Pathological lying	0.71	0.70	0.82	0.70	0.75	0.70
5 Conning/manipulative	0.90	0.74	0.95	0.77	0.92	0.75
6 Lack of remorse or guilt	1.28	0.70	1.10	0.72	1.21	0.72
7 Shallow affect	0.80	0.70	0.64	0.68	0.74	0.70
8 Callous/lack of empathy	1.09	0.71	0.92	0.72	1.02	0.72
9 Parasitic lifestyle	0.62	0.65	0.62	0.67	0.62	0.66
10 Poor anger control	1.36	0.71	1.49	0.67	1.41	0.70
11 Impersonal sexual behavior	0.79	0.82	0.79	0.83	0.79	0.82
12 Early behavior problems	1.01	0.81	0.77	0.81	0.92	0.82
13 Lacks goals	1.22	0.77	0.98	0.79	1.13	0.79
14 Impulsivity	1.26	0.65	1.37	0.58	1.30	0.63
15 Irresponsibility	1.31	0.67	1.33	0.65	1.32	0.66
16 Failure to accept responsibility	1.08	0.77	1.11	0.77	1.09	0.77
17 Unstable interpersonal relationships	0.70	0.72	0.98	0.77	0.81	0.75
18 Juvenile delinquency	1.65	0.52	1.43	0.66	1.57	0.59
19 Serious violations of conditional release	1.10	0.84	1.17	0.87	1.13	0.85
20 Criminal versatility	1.29	0.83	1.18	0.81	1.24	0.82
PCL: YV 20 items Total	20.73	7.33	20.13	7.49	20.50	7.39
PCL: YV 13 items Total	12.79	5.11	12.28	4.97	12.59	5.06

[a] $n = 614$.

[b] $n = 384$.

[c] $N = 1,007$.

Note. PCL: YV scores are presented in raw scale units (0 = does not apply; 1 = applies to a certain extent; 2 = definitely applies).

Table 3

Estimated item parameters for the generalized partial credit model (G-PCM) and corresponding factor loadings

Item	G-PCM estimated parameters					Factor loadings			
	a	β	δ_1	δ_2	Factor 1	Factor 2	Factor 3	h^2
Factor 1[a]								
1 Glibness/superficial charm	1.18	0.83	0.25	1.41	0.70			0.49
2 Grandiose sense of self-worth	1.20	0.48	−0.06	1.02	0.73			0.54
4 Pathological lying	0.59	0.67	−0.27	1.60	0.55			0.31
5 Conning/manipulative	0.70	0.18	−0.54	0.90	0.66			0.43
mean	*0.92*	*0.54*						
Factor 2[b]								
6 Lack of remorse or guilt	1.59	−0.37	−1.08	0.33		0.79		0.62
7 Shallow affect	0.63	0.68	−0.27	1.63		0.53		0.28
8 Callous/lack of empathy	1.15	−0.05	−0.84	0.75		0.81		0.65
16 Failure to accept responsibility for own actions	0.46	−0.25	−0.94	0.44		0.44		0.19
mean	*0.96*	*0.00*						
Factor 3[c]								
3 Need for stimulation	0.87	−0.45	−1.23	0.34			0.68	0.46
9 Parasitic lifestyle	0.42	1.33	0.11	2.55			0.43	0.19
13 Lacks goals	0.24	−0.55	−1.15	0.04			0.25	0.06
14 Impulsivity	0.52	−1.05	−2.54	0.45			0.44	0.20
15 Irresponsibility	0.64	−0.87	−1.94	0.20			0.50	0.25
mean	*0.54*	*−0.32*						

Note.

[a]-2 log Likelihood = 7311.43.

[b]-2 log Likelihood = 7535.68.

[c]-2 log Likelihood = 9424.42.

a = slope parameter. β = location parameter. δ_1, δ_2 = intersection parameters. h^2 = communality, indicating the variance accounted for by the factors.

Table 4

Differential item functioning for boys and girls

Item	α^a	Boys ($n = 614$)			Girls ($n = 384$)			β_b-β_g	Prob.
		β	δ_1	δ_2	β	δ_1	δ_2		
Factor 1^b									
1 Gilbness/superficial charm	1.17	0.85	0.27	1.43	0.90	0.32	1.48	0.05	.621
2 Grandiose sense of self-worth ^{e}	1.21	0.44	−0.10	0.98	0.64	0.10	1.18	0.20	.037
4 Pathological lying	0.59	0.79	−0.15	1.73	0.57	−0.38	1.51	−0.22	.088
5 Conning/manipulative	0.70	0.23	−0.49	0.95	0.20	−0.52	0.92	−0.03	.795
mean	*0.92*	*0.58*			*0.58*				
Factor 2^c									
6 Lack of remorse or guilt	1.61	−0.51	−1.21	0.20	−0.43	−1.13	0.28	0.08	.360
7 Shallow affect ^{e}	0.60	0.53	−0.45	1.51	0.75	−0.23	1.73	0.22	.088
8 Callous/lack of empathy	1.13	−0.18	−0.98	0.62	−0.11	−0.91	0.70	0.07	.436
16 Failure to accept responsibility for own actions ^{e}	0.47	−0.21	−0.90	0.48	−0.59	−1.28	0.10	−0.38	.007
mean	*0.95*	*−0.09*			*−0.09*				
Factor 3^d									
3 Need for stimulation	0.88	−0.47	−1.26	0.31	−0.54	−1.32	0.25	−0.07	.552
9 Parasitic lifestyle	0.41	1.34	0.11	2.56	1.20	−0.02	2.42	−0.14	.491
13 Lacks goals ^{e}	0.24	−0.95	−1.58	−0.32	−0.03	−0.67	0.60	0.92	< .0001
14 Impulsivity ^{e}	0.52	−0.89	−2.39	0.61	−1.44	−2.94	−0.14	−0.55	.001
15 Irresponsibility	0.63	−0.86	−1.93	0.22	−1.03	−2.10	0.05	0.17	.205
mean	*0.54*	*−0.37*			*−0.37*				

Note.

^{a} Slope parameter constrained to be the same between boys and girls.

^{b} -2 log Likelihood = 7300.38.

^{c} -2 log Likelihood = 7524.84.

^{d} -2 log Likelihood = 9392.77.

[e]Significant differential item functioning in the estimated location parameters (β).

a = slope parameter. β = location parameter. δ_1, δ_2 = intersection parameters.