

# Evaluating the Screening Version of the Hare Psychopathy Checklist—Revised (PCL:SV): An Item Response Theory Analysis

David J. Cooke and Christine Michie  
Glasgow Caledonian University

Stephen D. Hart  
Simon Fraser University

Robert D. Hare  
University of British Columbia

The Screening Version of the Psychopathy Checklist—Revised (PCL:SV; S. D. Hart, D. N. Cox, & R. D. Hare, 1995) was developed to complement the Psychopathy Checklist—Revised (PCL–R; R. D. Hare, 1991), and for use outside forensic settings. The PCL:SV takes less time to administer and requires less collateral information than the PCL–R. An item response theory approach was adopted to determine similarities in the structural properties of the 2 instruments and whether the PCL:SV could be regarded as a short form of the PCL–R. Eight of the 12 items in the PCL:SV were strongly parallel to their equivalent PCL–R items. Of the 4 items PCL:SV items which differed from their equivalent PCL–R items, all 4 were found to be equal or superior to their equivalent PCL–R items in terms of discrimination. The analyses confirmed previous results that the interpersonal and affective features of psychopathy have higher thresholds than do the impulsive and antisocial behavioral features; individuals have to be at a higher level of the psychopathic trait before the interpersonal and affective features become evident. The PCL:SV is an effective short form of the PCL–R.

The Hare Psychopathy Checklist—Revised (PCL–R; Hare, 1991) is a rating scale designed to measure traits of psychopathic personality disorder in forensic populations. It comprises 20 items, each reflecting a different symptom or characteristic of psychopathy (see Table 1). Items are defined in detail in the PCL–R manual and are rated on a 3-point scale (0 = *item doesn't apply*, 1 = *item applies somewhat*, 2 = *item definitely applies*). The items are summed to yield total scores, ranging from 0 to 40, that reflect the degree to which an individual resembles the prototypical psychopath. A cutoff score of 30 or greater is used to diagnose psychopathy. The items also can be summed to yield scores on two moderately correlated factors (Cooke, 1995; Hare et al., 1990; Harpur, Hakstian, & Hare, 1988; Harpur, Hare, & Hakstian, 1989). Factor 1 reflects the affective and interpersonal features of psychopathy and has been labeled the Selfish, Callous, and Remorseless Use of Others; Factor 2 reflects the social deviance features of psychopathy and has been labeled, Chronically Unstable and Antisocial lifestyle (Hare et al., 1990).

The PCL–R was developed originally for use in experimental psychopathology, and it has yielded a large and impressive body of

research (see Cooke, Forth, & Hare, 1998 for a review). However, more recently, the PCL–R has received considerable attention from forensic researchers because of its predictive validity with respect to criminal behavior, and in particular violent crime (for recent reviews, see Hart & Hare, 1996, 1997; Salekin, Rogers, & Sewell, 1996). For example, the PCL–R is related to failure on conditional release (e.g., Hart, Kropp, & Hare, 1988; Hemphill, Hare, & Wong, 1998; Serin, Peters, & Barbaree, 1990), violent recidivism (Harris, Rice, & Quinsey, 1993; Quinsey, Rice, & Harris, 1995; Serin, 1996; Serin & Amos, 1995), and poor treatment response (Ogloff, Wong, & Greenwood, 1990; Rice, Harris, & Cormier, 1992). Reviewers have commented favorably on the PCL–R's psychometric properties and its criterion- and construct-related validities (e.g., Fulero, 1995; Rogers, 1995; Stone, 1995).

One problem with the PCL–R is that administration requires access to detailed interview and case history information, including a criminal record. This means that the PCL–R is costly to administer in terms of time and effort, and it also may be inappropriate for use outside of forensic settings where criminal records often are unavailable or irrelevant (thus rendering several items unscorable). These considerations led to the development of the Screening Version of the PCL–R (PCL:SV; Hart, Cox, & Hare, 1995). The PCL:SV is a 12-item rating scale based directly on the PCL–R. The item descriptions in the PCL:SV manual are very brief relative to those for the PCL–R and require less detailed information to score. Most of the PCL:SV items were derived directly from a single PCL–R item by shortening and simplifying the item without losing its essential meaning. Thus, one would expect the corresponding PCL:SV and PCL–R items to be strongly parallel in psychometric terms. The remaining PCL:SV items were

---

David J. Cooke and Christine Michie, Department of Psychology, Glasgow Caledonian University; Stephen D. Hart, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada; Robert D. Hare, University of British Columbia, Vancouver, British Columbia, Canada.

Correspondence concerning this article should be addressed to David J. Cooke, The Douglas Inch Centre, 2 Woodside Terrace, Glasgow, G3 7UY, United Kingdom. Electronic mail may be sent to djcooke@rgardens.u-net.com.

Table 1  
Items in the Hare Psychopathy Checklist—Revised

Item	Factor loading
Glibness/superficial charm	1
Grandiose sense of self-worth	1
Need for stimulation/proneness to boredom	2
Pathological lying	1
Conning/manipulative	1
Lack of remorse or guilt	1
Shallow affect	1
Callous/lack of empathy	1
Parasitic lifestyle	2
Poor behavioral controls	2
Promiscuous sexual behavior	—
Early behavior problems	2
Lack of realistic, long-term goals	2
Impulsivity	2
Irresponsibility	2
Failure to accept responsibility for own actions	1
Many short-term marital relationships	—
Juvenile delinquency	2
Revocation of conditional release	2
Criminal versatility	—

Note. Factor 1 = Selfish, Callous, and Remorseless Use of Others; Factor 2 = Chronically Unstable and Antisocial Lifestyle. A dash indicates the item does not load on either factor.

derived by first collapsing, then shortening and simplifying, two PCL-R items that are highly similar in content. For example, PCL:SV Item 5 (lacks empathy) reflects a combination of PCL-R Items 8 (callous/lack of empathy) and 7 (shallow affect). PCL:SV items that reflect antisocial behavior also were modified so that scoring did not require access to a formal criminal record. Part 1 of the PCL:SV comprises 6 items that are intended to reflect Factor 1 of the PCL-R; the 6 items of Part 2 are intended to reflect Factor 2 of the PCL-R. Table 2 lists the 12 PCL:SV items as well as PCL-R item or items from which they were derived.

The PCL:SV items are rated on the same 3-point scale used for the PCL-R and are summed to yield total scores ranging from 0 to 24. A cutoff score of 18 or greater is used to diagnose psychopathy. The items can be summed to yield scores on two moderately correlated factors, isomorphic to those on the PCL-R. The first six PCL:SV items are labeled "Part 1" and correspond with Factor 1 of the PCL-R; the last six are labeled "Part 2" and correspond with Factor 2 of the PCL-R. Research presented in the PCL:SV manual suggests that it is a reliable and valid measure of psychopathy. Most important, it correlated highly ( $r > .80$ ) with the PCL-R in several samples, even when the scales were administered on separate occasions by independent raters.

### Evaluating the PCL-R Using Item Response Theory

Most evaluations of the PCL-R's psychometric or structural properties have been based on classical test theory. In general, these evaluations have concluded that the PCL-R seems appropriate for use with a wide range of forensic populations, including mentally disordered offenders, a variety of different ethnic and linguistic minorities in North America and offenders in several European nations (e.g., af Klinteberg, Humble, & Schalling, 1992; Cooke, 1995, 1998; Côté, 1990; Forth, Hart, & Hare, 1990; Haa-

pasalo & Pulkkinen, 1992; Hare et al., 1990; Hart & Hare, 1989; Kosson, Smith, & Newman, 1990; Raine, 1985; Wong, 1984). However, Cooke and Michie (1997, in press) have argued that item response theory has several advantages over classical test theory as a framework for evaluating the PCL-R, and particularly in evaluating its suitability for crosscultural research. Item response theory approaches to the analysis of test items and test functioning are distinct from, but complementary to, classical test theory methods. Item response theory models use mathematical functions to specify the relation between item responses and the latent trait postulated to determine those responses. The relations between item responses and latent trait can be expressed graphically as item characteristic curves (ICCs), also known as trace lines (Lazarsfeld, 1950). The advantages of item response theory include the following (Cooke & Michie, 1997, in press; Embretson, 1996; Nunnally & Bernstein, 1994; Steinberg & Thissen, 1996):

1. Item response theory parameters are invariant with respect to the samples from which they are generated but conditional on the underlying latent trait. Indeed, the parameters are independent of the test used to generate them: Thus, item response theory models offer the possibility of "test-free measurement" (Nunnally & Bernstein, 1994, p. 395). In contrast, standard classical test theory indexes of reliability and validity, such as corrected item-to-total correlations and coefficient alpha, are sensitive to variation across samples with respect to the version of the test used or the range of observed test scores.

2. Item response theory analyses determine the precision of a test's measurement at any value of the latent trait. Classical test theory, on the other hand, merely estimates precision of measurement at the mean trait level of the sample. Item or test information provide more data about the precision with which the underlying trait is being measured across the entire range of the trait.

3. Item response theory analyses can quantify the discriminative power of individual items and the overall test score at any value of the latent trait. If the test is diagnostic in nature, then item

Table 2  
Items in the Hare Psychopathy Checklist—Revised (PCL-R):  
Both Screening Version (PCL:SV) and the  
Corresponding PCL-R Items

PCL:SV item	Corresponding PCL-R items
Part 1	
Superficial	1
Grandiose	2
Deceitful	4, 5
Lacks remorse	6
Lacks empathy	7, 8
Doesn't accept responsibility	16
Part 2	
Impulsive	3, 14
Poor behavior controls	10
Lacks goals	9, 13
Irresponsible	15
Adolescent antisocial behavior	12, 18
Adult antisocial behavior	19, 20

Note. Two PCL-R items (many short-term marital relationships and promiscuous sexual behavior) were not used in the development of the PCL:SV. See Table 1 for full listing of the PCL-R items.

response theory analyses can be used to select items that provide maximum information at or around the diagnostic cutoff point. On the other hand, if the test is designed to measure trait strength, then item response theory analyses can be used to ensure that there are sufficient items to provide information at every level of the latent trait.

4. Item response theory analyses permit direct comparison of the performance of parallel items, that is, different items purporting to measure the same domain in the same sample, or the same item in different samples. Thus, item response theory is well-suited for determining whether item and test scores are invariant across forms (e.g., original vs. revision or translation, full vs. short form) and respondents (e.g., men vs. women, ethnic minority vs. ethnic majority). In the current study, the 20 PCL-R items and the 12 PCL:SV items are estimated simultaneously to ensure that all estimates of parameters and trait levels are on a common continuum of variation and so can validly be compared.

5. Item response theory methods can ensure that there is measurement invariance in the underlying trait being measured by the two instruments that are being compared. "Anchoring" procedures can be used to investigate whether the metrics of the underlying trait measured by the two scales are the same (Reise, Widaman, & Pugh, 1993). Thus, an individual whose score on the underlying latent trait is 1.5 when measured by the PCL-R should also have a score on the latent trait of 1.5 when measured by the PCL:SV. It is not possible to determine whether measurement invariance exists with classical test theory.

Cooke and Michie (1997) used item response theory methods to reanalyze the PCL-R normative data, which were collected primarily in North America. The ICCs of the PCL-R items indicated that they were all related to the latent trait in a meaningful way. Furthermore, together the items provided good discrimination across almost the entire range of the latent trait, and in particular, around the diagnostic cutoff. Finally, Cooke and Michie's analyses indicated that items from Factor 1 of the PCL-R were, in general, more central to the latent trait than were those from Factor 2: Factor 1 items carried more information, and they were also more likely to be discriminant at higher levels of the latent trait.

### Present Study

To date, the psychometric properties of the PCL:SV have been analyzed only using classical test theory methods (Hart et al., 1995). In the present study, we reanalyze normative data from the PCL:SV manual in an attempt to answer two basic questions: (a) Are the structural properties of the PCL:SV similar to those of the PCL-R, and (b) can the PCL:SV be considered a short form (or parallel form) of the PCL-R? In examining the structural properties, we considered three issues that can, as noted in the discussion above, be tackled using item response theory methods but not with classical test theory approaches. First, do similarly defined items from the different tests behave similarly? In particular, are they equally discriminating and are they positively endorsed in the same range of the underlying trait? Second, do both tests measure the same underlying trait in the same way; is there metric equivalence of the underlying trait measured using the two procedures? Third, are the tests equally informative, and does the point of maximum information vary between the tests?

## Method

### PCL:SV Ratings

We reanalyzed the normative data collected as part of the standardization of the PCL:SV. A full description of the standardization samples is provided in the PCL:SV manual. Briefly, there were a total of 586 participants from 11 different samples in Canada and the United States. The samples came from one of four settings: forensic nonpsychiatric (i.e., convicted prisoners not identified as mentally ill),  $N = 149$ ; forensic psychiatric (i.e., individuals charged with or convicted of offenses who were identified formally as mentally disordered and who were being treated either as inpatients or outpatients),  $N = 120$ ; civil psychiatric (i.e., individuals formally identified as mentally disordered, but with no current charges or convictions, and being treated as inpatients and outpatients),  $N = 217$ ; and civil nonpsychiatric (i.e., community residents and university students who are not currently identified as mentally disordered and with no current charges or convictions),  $N = 100$ . Participants ranged in age from 16 years to 82 years ( $M = 31$  years); 181 (31%) were female. The settings were selected to be representative of the settings where the instruments might be used; the participants were volunteers.

PCL:SV assessments were conducted by trained evaluators on the basis of an interview and a review of file information. To evaluate the concurrent validity of the PCL:SV, we assessed a subsample of 244 participants, using the PCL-R. The PCL-R assessments were conducted naive to PCL:SV ratings, that is, by independent raters on the basis of a new interview and review of file information about 1 week before or after the PCL:SV assessment.

The total sample, who were assessed by the PCL:SV, and the subsample, who were additionally assessed using the PCL-R, not only provided a wide range on the trait but also were broadly similar in the gender and the settings from which they were drawn (e.g., forensic inpatients and outpatients). It should be noted, however, that the PCL-R was not completed for the student samples, as many of the items could not be scored with such participants.

### Selection of Item Response Theory Model

The first and perhaps most important step in conducting an item response theory evaluation is the selection of a mathematical model from the range of models that has been developed for different measurement situations (Holland & Wainer, 1993; Hulin, Drasgow, & Parsons, 1983; Lord, 1980; Nunnally & Bernstein, 1994; Van der Linden & Hambleton, 1996). Two features of PCL:SV and PCL-R are important to consider here. First, the item response format is ordinal, comprising three ranked categories; and second, as assessments are conducted by trained raters, item scores are unlikely to be influenced by guessing. Consequently, the most appropriate model appears to be the three category version of Samejima's (1969) graded model (Cooke & Michie, 1997).

A thorough discussion of item response theory is beyond the scope of this article (see Nunnally & Bernstein, 1994; Thissen & Steinberg, 1988; Van der Linden & Hambleton, 1996; Wainer & Mislevy, 1990). Briefly, the latent trait that accounts for covariation among test items is denoted as  $\theta$ . According to the model selected, and assuming  $\theta$  is a unidimensional trait, the probabilities of receiving ratings of 0, 1, or 2 on item  $i$  are functions of  $\theta$ ,  $a$ ,  $b_1$ , and  $b_2$ .

$$P(0) = P(\text{Response} = 0|\theta) = 1 - \frac{1}{1 + \exp[-a(\theta - b_1)]}$$

$$P(1) = P(\text{Response} = 1|\theta)$$

$$= \frac{1}{1 + \exp[-a(\theta - b_1)]} - \frac{1}{1 + \exp[-a(\theta - b_2)]}$$

$$P(2) = P(\text{Response} = 2|\theta) = \frac{1}{1 + \exp[-a(\theta - b_2)]}$$

Plots of these functions, known as trace lines, for one of the PCL:SV items are shown in Figure 1.

The  $a$  parameter is an index of the discriminating power of the response; the larger the  $a$ , the more discriminating the item. The  $a$  parameter is the item response theory equivalent of the corrected item–total correlation, or  $r_{it}$ , in classical test theory. The position of the points of inflection are described by the threshold parameters  $b_i$ , which are indices of item difficulty (extremity, unpopularity);  $b_1$  for  $p(0)$  and  $b_2$  for  $p(2)$ . Items with large  $b_1$  parameters are endorsed positively only when  $\theta$  is relatively large, and increases in  $b_1$  result in the shifting of the ICC to the right. The  $b_1$  parameters are the item response theory equivalent of the proportion of correct responses, or  $p$ , in classical test theory CTT.

### Data Analyses

Our evaluation of the PCL:SV required two general steps. First, we constructed separate item response theory models for the PCL:SV and PCL–R. This required demonstrating that the PCL:SV and PCL–R were unidimensional scales, as our selected model assumes that  $\theta$  is unidimensional; this is variously referred to as the assumption of *homogeneity*, *local independence*, or *conditional independence*.<sup>1</sup> Also in this step, we tested the generality of the item response theory models across the gender of subjects.

In the second step, the one of primary interest in this article, we compared the item response theory models for the PCL:SV and PCL–R. At the item level, we examined the similarity of the ICCs for corresponding PCL:SV and PCL–R items (or item pairs). At the level of composite scores, we examined the relationship between the latent traits ( $\theta$ s) measured by the PCL:SV and PCL–R items. Next, we compared the test and item information functions for the PCL:SV and PCL–R. Finally, we determined the diagnostic efficiency of the PCL:SV with respect to the PCL–R.

All item response theory analyses were conducted using Multilog (Thissen, 1991), a well-regarded and widely used computer program (e.g., Nunnally & Bernstein, 1994). The model fitting procedure is an estimation minimalization (EM) algorithm developed by Bock and Aitken (1981).

## Results

### Unidimensionality

There is a lack of consensus concerning how best to assess dimensionality (Hattie, 1985). Traditionally, it has been evaluated using Cronbach's  $\alpha$ , mean  $r_{it}$ , or the relative proportion of variance accounted for by the first and second unrotated principal components. Using these criteria, we found the PCL:SV appeared to be unidimensional in the current sample of 586 participants: Cronbach's  $\alpha$  was .96, mean  $r_{it}$  was .66, and the ratio between the first and second unrotated principal components was about 5:1. Hare (1991) argued that the PCL–R also is essentially homogeneous, with a second-order factor being underpinned by two correlated first-order factors. Analysis of the PCL–R ratings in this data set supported Hare's findings: Cronbach's  $\alpha$  was .93, mean  $r_{it}$  was .60, and the ratio of the variance accounted for by the first and second unrotated principal components was about 6:1. In light of these results, we concluded that the PCL:SV and PCL–R were sufficiently unidimensional to permit their evaluation using item response theory methods. Confirmatory factor analysis has demonstrated that both instruments fit a hierarchical model with a highly saturated superordinate factor overpinning first-order factors (Cooke & Michie, 1998).

### Generality of Item Response Theory Models

Because of limited sample size, it was impossible to fit item response theory models to the data from each individual subsample in the data set. Therefore, we were unable to investigate whether the item response theory parameters were the same across settings. However, an attempt was made to develop separate PCL:SV models for men and women. By comparing these models, we would be able to determine whether there was evidence of differential item functioning (DIF). DIF would be suggested by the presence of significant differences in the value of item parameters across gender. In the PCL:SV normative data set, complete ratings were available for 405 men and 181 women. Using Multilog, we fitted a three category version of Samejima's (1969) graded model to these data by the method of maximum likelihood. The run for men was successful; however, the run for women failed to converge because a matrix became singular. This singularity was most probably a consequence of the limited range of  $\theta$  in the female sample, and the large number of parameters being estimated (36 item parameters) and the small sample size ( $n = 181$ ). To combat this problem, we compared the item parameters for men and women in batches of six items at a time: the first six items, the middle six items, and the last six items. We found significant differences ( $p < .05$ ) in the first of these three comparisons. On inspection, these differences appeared to be due to a large difference in the  $b_2$  parameter for Item 5, Lacks Empathy. As the observed differences are due to only one item parameter and may well have been the consequence of multiple-significance testing, we made the assumption—on the principle of parsimony—that there was no DIF across gender in this sample. However, the sample size was relatively small; it is possible that future analyses based on larger samples will find DIF. It was not possible to examine DIF across gender for the PCL–R, due to small sample size.

### Comparison of PCL:SV and PCL–R Items

Multilog was used to construct a graded model that contained the 12 PCL:SV items and the 20 PCL–R items. Estimating the parameters of all 32 items simultaneously ensures that we have a common metric, and thus, it is valid to compare item parameters or estimates of trait. The fitted parameters for this unconstrained model are presented in Table 3. To facilitate comparisons, corresponding PCL:SV and PCL–R items are aligned.

It is apparent from Table 3 that the parameters of corresponding items are similar in many instances. The method of generalized likelihood ratio testing (GLRT) was used to investigate the degree of similarity. According to the theory of GLRT, under certain conditions, imposing constraints on such a model, for example, constraining certain parameters to be equal, leads to an increase in the statistic  $G^2 = -2 \log$  likelihood. The increase is distributed as a  $\chi^2$  statistic with degrees of freedom equal to the number of constraints imposed.

A series of nested analyses was carried out. At each step in the

<sup>1</sup> This is not a strong assumption, however, as IRT models are relatively robust to departures from unidimensionality. Indeed, some IRT methods explicitly take into account and analyze multidimensionality (e.g., Steinberg & Thissen, 1996; Yen, 1993).

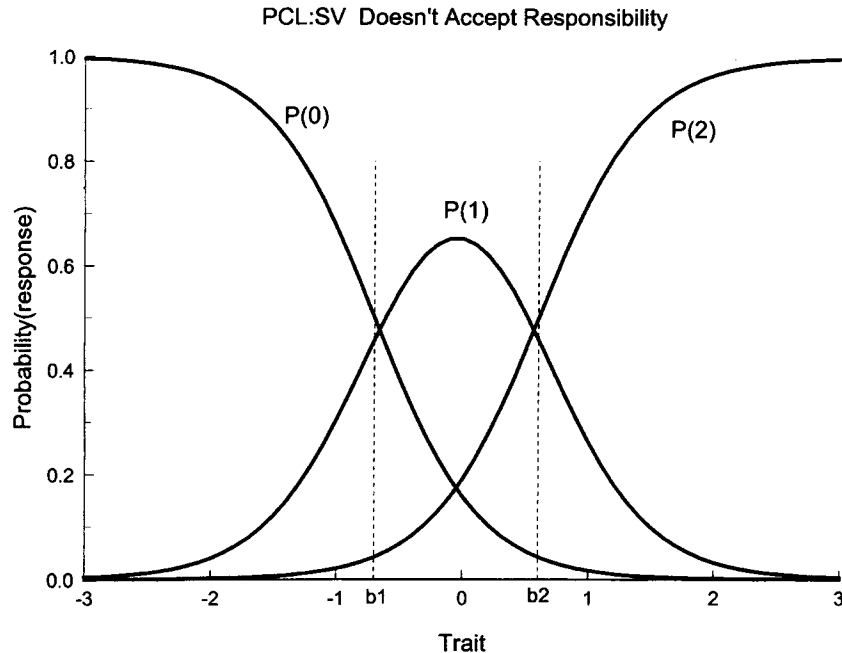


Figure 1. Example of item response theory curves. PCL:SV = Screening Version of the Psychopathy Checklist—Revised.

nested analyses, the parameters of an additional pair of corresponding items were constrained to be identical. The resulting increase in  $G^2$  over that for the unconstrained model was tested, and if not significant, then the item pair remained constrained in the next analysis. The analyses were carried out in four stages. At the first stage, item pairs in which a PCL:SV item corresponded with only one PCL-R item were constrained. At the second stage, pairs of PCL-R items that corresponded with one PCL:SV item

were constrained to be identical. At the third stage, the two PCL-R pairs that did not cause a significant increase in  $G^2$  were constrained to have identical parameters to the corresponding PCL:SV item, that is, all three items were constrained to have equal parameters. At the fourth stage, we investigated the four PCL-R pairs that did not have equivalent parameters. In these cases, the parameters of the PCL:SV item were constrained to be equal to the item parameters of each of the PCL-R items in turn. The sequence

Table 3

Comparison of Hare Psychopathy Checklist: Screening Version (PCL:SV) and Revised (PCL-R) Items: Full Model

PCL:SV				PCL-R			
Item	$a$	$b_1$	$b_2$	Item	$a$	$b_1$	$b_2$
Superficial	1.5	-0.1	1.6	Superficial	1.6	-0.2	1.5
Grandiose	1.6	-0.2	1.3	Grandiose	1.5	-0.3	1.4
Deceitful	1.7	-0.4	1.3	Conning/manipulative	1.8	-0.5	1.2
Lacks remorse	3.2	-0.5	0.6	Pathological lying	1.6	0.0	1.6
Lacks empathy	2.9	-0.5	1.1	Lack of remorse	2.4	-0.9	0.4
Doesn't accept responsibility	2.4	-0.7	0.6	Callous/lack of empathy	3.5	-0.5	0.6
Impulsive	2.0	-1.2	0.2	Shallow affect	2.1	-0.8	0.8
Poor behavior controls	1.6	-0.8	0.5	Failure to accept responsibility for own actions	1.7	-0.9	0.5
Lacks goals	1.6	-1.2	0.4	Impulsivity	2.4	-1.0	0.3
Irresponsible	2.3	-1.0	0.4	Need for stimulation	2.0	-1.1	0.4
Adolescent antisocial behavior	1.4	-0.8	0.8	Poor behavior controls	1.8	-0.6	0.6
Adult antisocial behavior	2.5	-0.8	0.3	Lack of goals	1.5	-1.0	0.6
				Parasitic lifestyle	2.1	-0.6	0.9
				Irresponsibility	2.6	-1.0	0.5
				Juvenile delinquency	1.3	0.0	1.4
				Early behavior problems	1.5	0.1	1.4
				Revocation of conditional release	1.6	-0.4	0.8
				Criminal versatility	2.1	0.2	0.8

Note.  $a$  is the slope of the logistic curve at the point of inflection;  $b_1$  and  $b_2$  are the thresholds for  $P_0$  and  $P_2$ , respectively.

of the analyses and the associated statistics are presented in Table 4. The item parameters of the resulting model with the constraints that did not result in significant increases in  $G^2$  are given in Table 5. The model fits the data well, predicting the observed pattern of responses for each item within 1%. In Table 5, corresponding items that have been constrained to have identical parameters are shown in boldface type.

In sum, the nested analyses suggested that 8 of 12 PCL:SV items (Items 1, 2, 3, 6, 7, 8, 9, and 10) had parameters identical to the PCL-R items from which they were derived. Thus, these items can be considered strongly parallel to their PCL-R counterparts. The remaining 4 PCL:SV items (Items 4, 5, 11, and 12) had parameters that were significantly different from the corresponding PCL-R items. We decided to focus on these nonparallel items to determine whether their inclusion in the PCL:SV was problematic.

GLRT indicated only that item parameters were different; it did not reveal how the items differed. Inspection of the parameters for the 4 nonparallel items (see Table 5) indicated that PCL:SV Item 4 (lacks remorse) was more discriminating and had higher thresholds than its corresponding PCL-R item. PCL:SV Item 5 (lacks empathy) was more discriminating than one member of the corresponding PCL-R item pair, but less discriminating than the other. Finally, PCL:SV Items 11 (adolescent antisocial behavior) and 12 (adult antisocial behavior) were equally as discriminating as the members of their corresponding PCL-R item pairs and had lower thresholds. These findings indicate that the four nonparallel PCL:SV items are equal or superior to their PCL-R counterparts with respect to discrimination, and 3 of 4 had lower thresholds; there is no evidence from this analysis that the nonparallel items should be dropped from the PCL:SV.

### Comparison of Part 1 and Part 2 Items

Items from Part 1 of the PCL:SV had, on average, higher  $b_1$  and  $b_2$  parameters than items from Part 2,  $U(6, 6) = 0.0, p < .001$ ;  $U(6, 6) = 2.0, p = .01$ ; there was no difference between items from Parts 1 and 2 with respect to the  $a$  parameter,  $U(6, 6) = 12.5, ns$ . These findings are consistent with earlier analyses of the PCL-R (Cooke & Michie, 1997, in press), which found that items loading on Factor 1 had a higher threshold than did items loading on Factor 2. Taken together, the findings suggest that interpersonal and affective features of psychopathy have a higher threshold than do impulsive and antisocial behavioral features; however, the interpersonal and affective features are no more discriminating.

### Comparison of PCL:SV and PCL-R Latent Traits

For 244 participants, there were four available estimates of the latent trait underlying the scales:  $\theta_{PCL:SV}$ , the estimate from the item response theory analysis of PCL:SV items;  $T_{PCL:SV}$ , the total score (sum of item scores) from the PCL:SV;  $\theta_{PCL-R}$ , the estimate from the item response theory analysis of PCL-R items; and  $T_{PCL-R}$ , the total score (sum of item scores) from the PCL-R. The correlations among these four estimates were very high (all  $r \geq .89$ ), and the associations were highly linear.

If the PCL:SV and PCL-R both measure the same latent trait, then the values of  $\theta_{PCL:SV}$  and  $\theta_{PCL-R}$  for a given participant should be the same within sampling error. A  $t$  value was calculated for each individual. The mean of these values was  $t = 4.93$  with 243 degrees of freedom  $p < .001$ , which is significant suggesting there is some difference in the traits measured by the PCL:SV and PCL-R items. To investigate this difference further, we drew a regression line and 95% confidence limits on a scatterplot of the

Table 4  
Series of Analyses to Reach Constrained Model

Analysis (i) + constraints imposed	$G_i^2$	$G_i^2 - G_o^2$	df	Probability
None	8,807.0			
Equate equivalent PCL:SV and PCL-R items				
Superficial	8,807.0	0.0	3	1.0
Analysis 1 + Grandiose	8,807.5	0.5	6	.998
Analysis 2 + Poor behavior controls	8,813.5	6.5	9	.689
Analysis 3 + Irresponsible	8,817.0	10.0	12	.616
Analysis 4 + Doesn't accept responsibility	8,829.8	22.8	15	.885 $\times 10^{-1}$
Analysis 5 + Lacks remorse	8,848.2	41.2	18	.143 $\times 10^{-3}$
Equate PCL-R items				
Analysis 5 + Juvenile delinquency & early behavior problems	8,831.5	24.5	18	.139
Analysis 7 + Need for stimulation & impulsivity	8,833.5	26.5	21	.188
Analysis 8 + Conning & pathological lying	8,848.0	41.0	24	.167 $\times 10^{-1}$
Analysis 8 + Revocation & criminal versatility	8,848.8	41.8	24	.136 $\times 10^{-1}$
Equate PCL:SV item and 2 PCL-R items				
Analysis 8 + Adolescent antisocial behavior	8,898.1	91.1	24	.947 $\times 10^{-9}$
Analysis 8 + Impulsivity	8,836.0	29.0	24	.220
Equate PCL:SV item and 1st PCL-R item				
Analysis 12 + Manipulative	8,837.2	30.2	27	.305
Analysis 13 + Lacks goals	8,846.4	39.4	30	.117

Note.  $df$  = total number of item parameters constrained to be equal;  $G_i^2$  is  $G^2$  (i.e.,  $-2 \log$  likelihood) for the  $i$ th analysis;  $G_o^2$  is  $G^2$  for the base (i.e., unconstrained) analysis. PCL:SV = Hare Psychopathy Checklist: Screening Version; PCL-R = Hare Psychopathy Checklist—Revised.

Table 5

Comparison of Hare Psychopathy Checklist: Screening Version (PCL:SV) and Revised (PCL-R) Items: Constrained Model

PCL:SV				PCL-R			
Item	<i>a</i>	<i>b</i> <sub>1</sub>	<i>b</i> <sub>2</sub>	Item	<i>a</i>	<i>b</i> <sub>1</sub>	<i>b</i> <sub>2</sub>
<b>Superficial</b>	<b>1.5</b>	<b>-0.2</b>	<b>1.6</b>	<b>Superficial</b>	<b>1.5</b>	<b>-0.2</b>	<b>1.6</b>
<b>Grandiose</b>	<b>1.6</b>	<b>-0.2</b>	<b>1.3</b>	<b>Grandiose</b>	<b>1.6</b>	<b>-0.2</b>	<b>1.3</b>
<b>Deceitful</b>	<b>1.7</b>	<b>-0.4</b>	<b>1.2</b>	<b>Conning/manipulative</b>	<b>1.7</b>	<b>-0.4</b>	<b>1.2</b>
Lacks remorse	3.1	-0.5	0.6	Pathological lying	1.7	0.0	1.5
Lacks empathy	2.8	-0.5	1.1	Lack of remorse	2.5	-0.9	0.3
<b>Doesn't accept responsibility</b>	<b>2.2</b>	<b>-0.7</b>	<b>0.6</b>	Callous/lack of empathy	3.6	-0.5	0.6
<b>Impulsive</b>	<b>2.0</b>	<b>-1.1</b>	<b>0.3</b>	Shallow affect	2.1	-0.8	0.7
<b>Poor behavior controls</b>	<b>1.6</b>	<b>-0.8</b>	<b>0.5</b>	<b>Failure to accept responsibility</b>	<b>2.2</b>	<b>-0.7</b>	<b>0.6</b>
<b>Lacks goals</b>	<b>1.5</b>	<b>-1.2</b>	<b>0.4</b>	<b>Impulsivity</b>	<b>2.0</b>	<b>-1.1</b>	<b>0.3</b>
<b>Irresponsible</b>	<b>2.4</b>	<b>-1.0</b>	<b>0.4</b>	<b>Need for stimulation</b>	<b>2.0</b>	<b>-1.1</b>	<b>0.3</b>
Adolescent antisocial behavior	1.4	-0.8	0.8	<b>Poor behavior controls</b>	<b>1.6</b>	<b>-0.8</b>	<b>0.5</b>
Adult antisocial behavior	2.5	-0.8	0.3	<b>Lack of goals</b>	<b>1.5</b>	<b>-1.2</b>	<b>0.4</b>
				Parasitic lifestyle	2.1	-0.6	0.9
				<b>Irresponsibility</b>	<b>2.4</b>	<b>-1.0</b>	<b>0.4</b>
				<b>Juvenile delinquency</b>	<b>1.4</b>	<b>0.0</b>	<b>1.4</b>
				<b>Early behavior problems</b>	<b>1.4</b>	<b>0.0</b>	<b>1.4</b>
				Revocation of conditional release	1.6	-0.4	0.8
				Criminal versatility	2.0	0.1	0.9

Note. Items that have been constrained to have identical parameters are in boldface. *a* is the slope of the logistic curve at the point of inflection; *b*<sub>1</sub> and *b*<sub>2</sub> are the thresholds for *P*<sub>0</sub> and *P*<sub>2</sub>, respectively.

estimates of trait from the PCL:SV items against the PCL-R items as shown in Figure 2.

The plot demonstrates that the relationship between the two estimates of trait is well described by the linear relationship. This means that the estimate of trait from the PCL:SV items is a good predictor of the estimate of trait from the PCL-R items and that the prediction is equally good across the entire range of the trait.

According to regression analysis, the relationship between  $\theta_{PCL:SV}$  and  $\theta_{PCL-R}$  can be expressed as follows:

$$\theta_{PCL:SV} = 0.130 + 0.963\theta_{PCL-R} \quad (1)$$

The regression analysis indicated that the association between the two latent traits was highly significant,  $F(1, 242) = 1,100.0$ ,  $r = .91$ ,  $p < .001$ . Although the association was linear in nature,

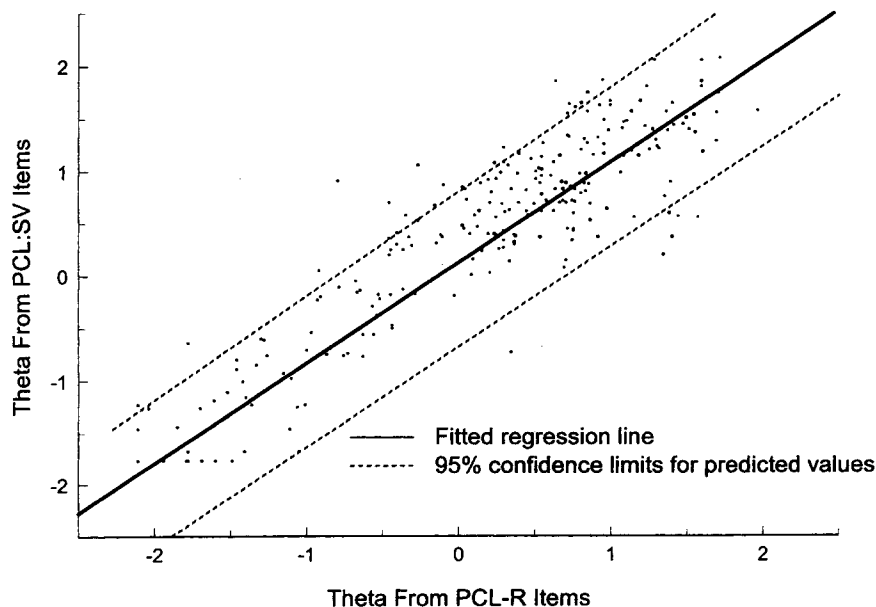


Figure 2. Scatter plot of theta values from Screening Version of the Psychopathy Checklist—Revised (PCL:SV) items versus theta values from PCL-R items. The fitted regression line with 95% confidence limits for prediction is plotted.

Table 6  
Information Functions at Various Levels of the Trait

Test	Information at various levels of $\theta$								
	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
PCL:SV	3.0	5.2	9.0	13.0	13.6	13.1	12.7	9.6	5.9
PCL-R	4.8	8.5	12.8	15.4	15.8	16.6	15.8	11.9	7.9
PCL-R/PCL:SV	1.6	1.6	1.4	1.2	1.2	1.3	1.2	1.2	1.3

Note. PCL:SV = Hare Psychopathy Checklist: Screening Version; PCL-R = Hare Psychopathy Checklist—Revised.

note that the intercept is significantly different from 0,  $t(242) = 5.0, p < .001$ . The slope does not differ significantly from 1,  $t(242) = 1.28, ns$ . These findings suggest that the PCL:SV measures a trait identical to that measured by the PCL-R with the addition of a small constant. However, the small magnitude of the difference and the high proportion of shared variance suggest the difference between the scales is not psychologically meaningful. Thus, it appears that the PCL:SV and PCL-R are metrically equivalent, but that they have different origins (by approximately 0.1 standardized units, i.e., equivalent to 1 in a PCL-R total score). Put another way, the two tests have the same units of measurement but slightly different zero points.<sup>2</sup> In fact, as the estimates of trait are discrete, the difference may be due to the fact that the possible estimates of trait from the PCL:SV items are different from those from the PCL-R items. We conclude that the PCL:SV can be considered a good measure of the intensity of the trait that previously has been estimated using the PCL-R.

#### Comparison of PCL:SV and PCL-R Information Functions

The information provided by a test or item is the inverse of the square of its standard error of estimate (Nunnally & Bernstein, 1994). Test or item information can be considered the item response theory equivalent of reliability in classical test theory. However, unlike CTT indexes of reliability, information functions provide estimates concerning precision of measurement at various levels of the latent trait. For diagnostic tests such as the PCL:SV and PCL-R, information should be a maximum around the diagnostic cutoff. Test information at various levels of  $\theta$  is presented in Table 6.

As would be expected from the difference in the length of the tests, the PCL-R provides more information than does the PCL:SV at all levels of the latent trait. Around the diagnostic cutoff (approximately  $\theta = 1.0$  on both tests; see below), the ratio of information is about 1.2:1. Thus, an increase in test length of 66% results in only a 20% increase in precision of measurement. For low levels of the trait, the ratio is much higher. This latter finding suggests that the PCL-R yields more precise measurements than does the PCL:SV at low levels of  $\theta$ , a finding that is somewhat surprising given that the PCL:SV was intended for use in nonforensic populations (where, presumably, levels of  $\theta$  are quite low), whereas the PCL-R was not.<sup>3</sup>

Item information functions are presented in Table 7. The PCL:SV and PCL-R items that had identical parameters also had identical information functions; thus, information functions for

PCL-R items are presented only for the four nonparallel PCL:SV items.

Inspection of Table 7 reveals that PCL:SV Part 1 items provided the most information at the highest levels of the latent trait, whereas Part 2 items provided the most information at low levels of the trait. Among Part 1 items, Items 4 (lacks remorse) and 5 (lacks empathy) provided the greatest information. Among Part 2 items, Items 12 (adult antisocial behavior) and 10 (irresponsible) provided the most information.

There are some noteworthy differences between the information functions of corresponding PCL:SV and PCL-R items; some of these differences are illustrated in Figure 3. For example, PCL:SV Item 5 (lacks empathy) generally carried more information than one of the PCL-R items from which it was derived, that is, shallow affect, but less information than the other item from which it was derived, that is, callous/lack of empathy. However, around the diagnostic cutoff (i.e., around 1.0 the information provided by items lacks empathy and callous/lack of empathy were essentially identical). The PCL:SV item, adolescent antisocial behavior, performed no better or no worse than the equivalent PCL-R items (i.e., juvenile delinquency and early behavior problems). By way of contrast, the PCL:SV item, adult antisocial behavior, provided substantially more information than either of the two PCL-R items from which it was derived (i.e., criminal versatility and revocation of conditional release).

Interesting to note, these information analyses provided strong support for retaining the nonparallel items in the PCL:SV. Three of the four PCL:SV items that carried the greatest information were nonparallel items; and all four nonparallel items carried as much information as did their PCL-R counterparts (albeit at different trait levels).

#### Diagnostic Efficiency of the PCL:SV

On the PCL-R, a cutoff score of  $\geq 30$  is used to classify individuals as psychopaths (Hare, 1991). Many researchers also consider those scoring  $\geq 21$  (but below 30) to constitute a group

<sup>2</sup> The relation between the PCL:SV and PCL-R may be considered analogous to that between the Celsius and Kelvin temperature scales.

<sup>3</sup> It should be kept in mind, however, that some PCL-R items are scored on the basis of formal criminal records. Thus, even though the PCL-R appears to have good precision at low levels of the latent trait in these analyses, for practical reasons it may be difficult or impossible to administer outside of forensic settings.



Table 7  
Item Information Functions

Item	Information $\times$ 100 at various levels of $\theta$								
	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
Superficial	13	24	40	55	62	61	61	62	53
Grandiose	14	27	45	62	69	68	69	63	47
Deceitful	17	35	60	78	78	76	80	73	50
PCL-R pathological lying	10	21	39	61	74	74	74	74	61
Lacks remorse	10	42	147	249	200	243	184	60	14
PCL-R lack of remorse or guilt	38	99	162	153	158	154	84	30	9
Lacks empathy	11	43	127	202	148	141	201	138	49
PCL-R callous/lack of empathy	6	36	171	337	240	332	200	45	8
PCL-R shallow affect	30	68	111	116	105	117	106	62	26
Doesn't accept responsibility	27	63	108	122	117	122	96	50	20
Impulsive	52	91	109	103	108	99	63	29	12
Poor behavior controls	27	46	64	73	74	70	55	35	19
Lacks goals	41	56	63	62	63	61	49	32	18
PCL-R parasitic lifestyle	22	52	93	111	100	105	108	76	37
Irresponsible	44	102	147	131	135	144	91	37	12
Adolescent antisocial behavior	26	39	50	55	55	55	50	39	26
PCL-R juvenile delinquency and early behavior problems	10	18	30	42	52	55	55	51	41
Adult antisocial behavior	29	79	149	162	162	150	80	29	9
PCL-R revocation of conditional release	17	31	50	66	72	71	65	48	29
PCL-R criminal versatility	5	14	34	70	109	122	109	72	35

Note. PCL-R = Hare Psychopathy Checklist—Revised.

with mixed or moderate psychopathic features (i.e., possible psychopaths). The PCL:SV manual recommends a cutoff of  $\geq 18$  to classify individuals as psychopaths and  $\geq 13$  to classify them as possibly psychopathic and in need of further evaluation with the 20-item PCL-R (Hart et al., 1995). Given the strong linear relationships between latent trait and test score on the PCL:SV and PCL-R, it is possible to check the appropriateness of the recommended PCL:SV cutoff, using Equation 1 and the following regression equations:

$$\theta_{PCL-R} = -1.71 + 0.092T_{PCL-R} \quad (2)$$

$$T_{PCL:SV} = 12.30 + 6.70\theta_{PCL:SV} \quad (3)$$

According to these equations, a cutoff of  $\geq 30$  on the PCL-R corresponds with  $\geq 20$  on the PCL:SV, and a cutoff of  $\geq 21$  on the PCL-R corresponds with  $\geq 15$  on the PCL:SV. Thus, according to these analyses, the recommended PCL:SV cutoffs both are about 2 points too low. However, this is not particularly problematic if the PCL:SV is used as a true screening test, as the probability of false-negative errors is very low and as false-positive errors would be detected in follow-up evaluation with the PCL-R.<sup>4</sup>

### Discussion

With respect to the two general questions posed in the Introduction, the item response theory analyses conducted here suggest that (a) the PCL:SV has structural properties very similar to those of the PCL-R, and, (b) the PCL:SV can be considered a short or parallel form of the PCL-R. It is perhaps remarkable, given the differences in assessment method and scoring criteria, that 8 of 12 PCL:SV items had ICC parameters identical to their corresponding PCL-R items. Furthermore, the other 4 PCL:SV items had parameters that made them equally useful or even more useful than their

corresponding PCL-R items; this may be due to the fact that the revision process made these PCL:SV items more general in scope. PCL:SV total scores were so strongly and linearly related to PCL-R total scores that the scales can be considered metrically equivalent measures of the same psychological construct. The theta-theta plot revealed that the two versions of the test are equivalent throughout the entire range of scores; this is particularly important given that the PCL:SV is a screening instrument. Note that although these item response theory analyses confirmed some previous findings based on classical test theory, they also examined some highly specific issues that could not have been addressed at all using classical test theory. This study, then, is a good illustration of the utility of item response theory for solving problems in applied measurement.

Some of the analyses also yielded findings of theoretical interest. One such finding concerned the parameters and information functions of items from Parts 1 and 2 of the PCL:SV. First, Part 1 items were as discriminating as Part 2 items, but had higher thresholds. This indicates that the impulsive and antisocial behavioral features of psychopathy were expressed at relatively low

<sup>4</sup> A more comprehensive evaluation of the PCL:SV cutoffs would require a larger sample size and statistical analyses beyond the scope of IRT, such as Receiver Operating Characteristic (ROC) analyses (e.g., Mossman, 1994). A preliminary ROC analysis using the present data suggested that a cutoff of  $\geq 19$  maximized overall predictive accuracy, yielding a true positive rate of 81% and a false positive rate of 15%; a cutoff of  $\geq 18$  was only slightly worse, with a true positive rate of 81% and a false positive rate of 20%. False negative errors were minimized by a cutoff of  $\geq 12$ , which yielded a false negative rate of 0% and a false positive rate of 60%. Given these findings, it may be premature to reject the recommended PCL:SV cutoffs at the present time.

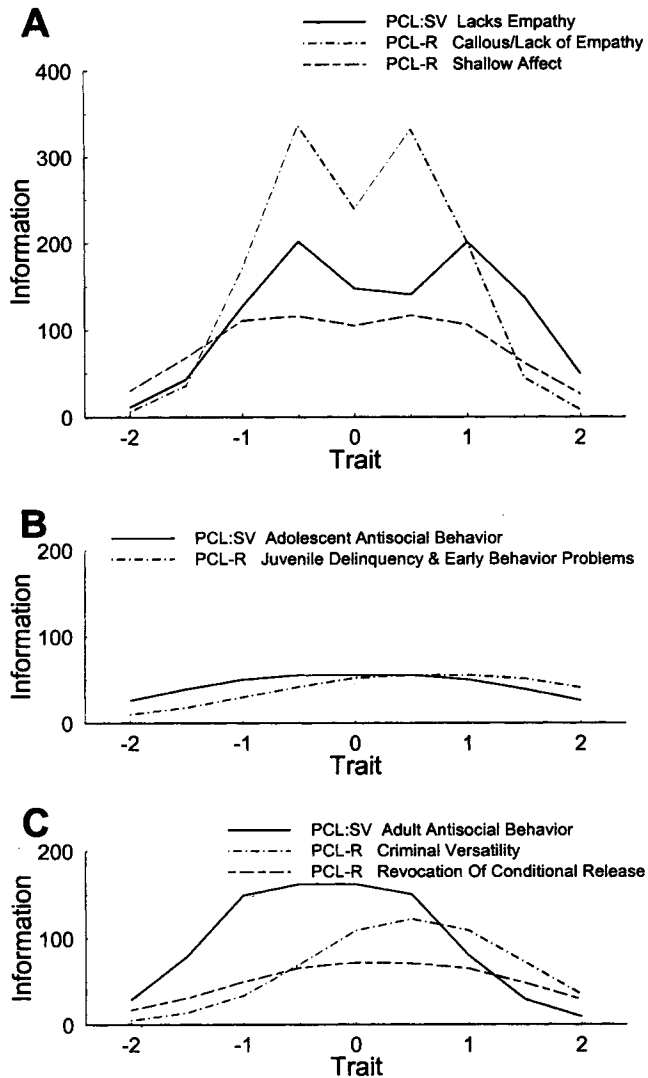


Figure 3. Information functions of selected items. (A) Information functions for PCL:SV item and the two PCL-R items from which it was derived. (B) Information functions for PCL:SV and PCL-R items that contribute relatively little to the trait estimates. (C) Information functions for PCL:SV items and two PCL-R items illustrating the improvement in information provided by the PCL:SV item. PCL:SV = Screening Version of Psychopathy Checklist—Revised; PCL-R = Psychopathy Checklist—Revised.

levels of the trait, and that the interpersonal and affective features of psychopathy were more central to the construct of psychopathy. Given that the interpersonal and affective features also are crucial to clinical conceptions of psychopathy, there seems to be strong grounds to include them in diagnostic criteria for psychopathy. Yet, they are grossly underrepresented in the criteria for antisocial personality disorder that appear in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*; American Psychiatric Association, 1994; see Hart & Hare, 1995).<sup>5</sup>

Second, Part 1 items carried as much information as did items from Part 2, suggesting that the interpersonal and affective features of psychopathy were assessed as reliably as the impulsive and

antisocial behavioral features. Similar results were reported by the *DSM-IV* antisocial personality disorder field trial (Widiger et al., 1996) and appear to nullify one of the primary arguments that has been used to keep the content of the diagnostic criteria for antisocial personality disorder so focused on behavioral features (Hart & Hare, 1995).

Despite the power of item response theory analyses, it is important to attempt to replicate the current findings. In particular, this study was unable to provide a clear answer concerning the generality of the derived item response theory model across gender; a larger sample of women is needed for this task. Also, larger data sets are required to determine whether the item response theory model will generalize across specific populations, such as civil psychiatric patients, college students, and so forth. Item response theory analyses of data collected that were collected in the *DSM-IV* antisocial personality disorder field trial (Widiger et al., 1996) are being carried out at present. These analyses will illustrate whether some of these findings generalize to other methods of measuring this important trait.

<sup>5</sup> The accompanying text in the *DSM-IV* discusses these interpersonal and affective features at some length as "associated features" of antisocial personality disorder.

## References

- af Klinteberg, B., Humble, K., & Schalling, D. (1992). Personality and psychopathy of males with a history of early criminal behaviour. *European Journal of Personality*, 6, 245-266.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Applications of an EM algorithm. *Psychometrika*, 46, 443-459.
- Cooke, D. J. (1995). Psychopathic disturbance in the Scottish prison population: The cross-cultural generalisability of the Hare psychopathy checklist. *Psychology, Crime, and Law*, 2, 101-108.
- Cooke, D. J. (1998). Cross-cultural aspects of psychopathy. In D. J. Cooke, A. E. Forth, & R. D. Hare (Eds.), *Psychopathy: Theory, research and implications for society*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Cooke, D. J., Forth, A. E., & Hare, R. D. (1998) (Eds.). *Psychopathy: Theory, Research, and Implications for Society*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Cooke, D. J., & Michie, C. (1997). An Item Response Theory evaluation of Hare's Psychopathy Checklist. *Psychological Assessment*, 9, 3-14.
- Cooke, D. J., & Michie, C. (1998). *Towards a hierarchical model of psychopathy*. Unpublished manuscript.
- Cooke, D. J., & Michie, C. (in press). Psychopathy across cultures: An item response theory comparison of Hare's Psychopathy Checklist—Revised. *Journal of Abnormal Psychology*.
- Côté, G. (1990). Interrater reliability and interrater agreement with the French version of Hare's Psychopathy Checklist (Abstract). *Canadian Psychology*, 31, 391.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8, 341-349.
- Forth, A. E., Hart, S. D., & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment*, 2, 342-344.
- Fulero, S. M. (1995). Review of the Hare Psychopathy Checklist—Revised. In Anonymous, *Twelve mental measurement yearbook* (pp. 453-454). Lincoln, NE: Buros Institute.

- Haapasalo, J., & Pulkkinen, L. (1992). The Psychopathy Checklist and non-violent offender groups. *Criminal Behaviour and Mental Health*, 2, 315–328.
- Hare, R. D. (1991). *Manual for the Revised Psychopathy Checklist* (1st ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The Revised Psychopathy Checklist: Descriptive statistics, reliability, and factor structure. *Psychological Assessment*, 2, 338–341.
- Harpur, T. J., Hakstian, R., & Hare, R. D. (1988). Factor structure of the psychopathy checklist. *Journal of Consulting and Clinical Psychology*, 56, 741–747.
- Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 6–17.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315–335.
- Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version* (1st ed.). Toronto, Ontario, Canada: Multi-Health Systems.
- Hart, S. D., & Hare, R. D. (1989). The discriminant validity of the Psychopathy Checklist in a forensic psychiatric population. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 211–218.
- Hart, S. D., & Hare, R. D. (1995). Commentary on antisocial personality disorder: The *DSM-IV* field trial. In W. J. Livesley (Ed.), *The DSM-IV personality disorders* (pp. 127–134). New York: Guilford Press.
- Hart, S. D., & Hare, R. D. (1996). Psychopathy and risk assessment. Psychopathy and antisocial personality disorder. *Current Opinion in Psychiatry*, 9, 129–132.
- Hart, S. D., & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. M. Stoff, J. Breiling, & J. Maser (Eds.), *The handbook of antisocial behavior* (pp. 22–35). New York: Wiley.
- Hart, S. D., Kropp, P. R., & Hare, R. D. (1988). The performance of male psychopaths following conditional release from prison. *Journal of Consulting and Clinical Psychology*, 56, 227–232.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139–164.
- Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, 3, 139–170.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning* (1st ed.). Hillsdale, NJ: Erlbaum.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Applications of IRT to language translations. In Anonymous, *Item Response Theory: Application to psychological measurement* (pp. 185–209). Homewood, IL: Dow Jones-Irwin.
- Kosson, D. S., Smith, S. S., & Newman, J. P. (1990). Evaluating the construct validity of psychopathy in Black and White male inmates: Three preliminary studies. *Journal of Abnormal Psychology*, 99, 250–259.
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, & J. A. Clausen (Eds.), *Measurement and prediction*. (pp. 362–412). New York: Wiley.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems* (1st ed.). Hillsdale, NJ: Erlbaum.
- Mossman, D. (1994). Further comments on portraying the accuracy of violence predictions. *Law and Human Behavior*, 18, 587–596.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Ogloff, J. R., Wong, S., & Greenwood, A. (1990). Treating adult psychopaths in a therapeutic community program within a correctional setting. *Behavioral Sciences and the Law*, 8, 81–90.
- Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10, 85–105.
- Raine, A. (1985). A psychometric assessment of Hare's checklist for psychopathy on an English prison population. *British Journal of Clinical Psychology*, 24, 247–258.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, 114, 552–566.
- Rice, M. E., Harris, G. T., & Cormier, C. A. (1992). An evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law and Human Behavior*, 16, 399–412.
- Rogers, R. (1995). *Diagnostic and structured interviewing: A handbook for psychologists*. Odessa, FL: Psychological Assessment Resources.
- Salekin, R., Rogers, R., & Sewell, K. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist—Revised: Predictive validity of dangerousness. *Clinical Psychology: Science & Practice*, 3, 203–215.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monographs*, 34 (4, Pt. 2, Whole No. 17).
- Serin, R. C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior*, 20, 207–217.
- Serin, R. C., & Amos, N. L. (1995). The role of psychopathy in the assessment of dangerousness. *International Journal of Law and Psychiatry*, 18, 231–238.
- Serin, R. C., Peters, R. D., & Barbaree, H. E. (1990). Predictors of psychopathy and release outcome in a criminal population. *Psychological Assessment*, 2, 419–422.
- Steinberg, L., & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods*, 1, 81–97.
- Stone, G. L. (1995). Review of the Hare Psychopathy Checklist—Revised. In J. C. Conoley & J. C. Impara (Eds.), *Twelfth mental measurement yearbook* (pp. 454–455). Lincoln, NE: Buros Institute.
- Thissen, D. (1991). *Multilog user's guide (Version 6)* (1st ed.). Mooresville, IN: Scientific Software.
- Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin*, 104, 385–395.
- Van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. New York: Springer.
- Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration, and proficiency estimation. In H. Wainer, N. Dorans, R. Flaugher, B. Green, R. Mislevy, L. Steinberg, & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (pp. 65–102). Hillsdale, NJ: Erlbaum.
- Widiger, T. A., Cadoret, R., Hare, R. D., Robins, L. N., Rutherford, M. J., Zhanarini, M., Alterman, A. I., Apple, M., Corbitt, E., Forth, A. E., Hart, S. D., Kultermann, J., & Woody, G. (1996). *DSM-IV* antisocial personality disorder field trial. *Journal of Abnormal Psychology*, 105, 3–16.
- Wong, S. (1984). *Criminal and institutional behaviours of psychopaths* (1st ed.). Ottawa, Ontario, Canada: Ministry of the Solicitor-General of Canada.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–214.

Received August 4, 1997

Revision received September 15, 1998

Accepted October 8, 1998 ■