Critical evaluation of psychopathy measurement (PCL-R and SRP-III/SF) and recommendations for future research

Daniel Boduszek

Agata Debowska

Paper accepted for publication in *Journal of Criminal Justice*

Author Note

Daniel Boduszek, PhD, 1) Department of Behavioural and Social Sciences, University of Huddersfield, Huddersfield, Queensgate, HD1 3DH, UK;

2) SWPS University of Social Sciences and Humanities, Psychology Department, Katowice, Poland.

Agata Debowska, PhD, Department of Psychology, University of Chester, Chester, Parkgate Road, CH1 4BJ, UK

This research was not supported by any research grant.

Correspondence concerning this article should be addressed to Daniel Boduszek, PhD, University of Huddersfield, Department of Behavioural and Social Sciences, Queensgate, Huddersfield, HD1 3DH, UK. Email: d.boduszek@hud.ac.uk
Abstract

**Purpose:** The purpose of this paper was to review, summarize, and critically engage with the most recent findings into the dimensionality of the PCL-R, SRP-III, and SRP-SF. Another objective was to provide a set of directions for future research.

**Methods:** A search in PubMed, PsychInfo, Scopus, Web of Science, Science Direct, and Google Scholar was performed. Twenty-one studies examining the dimensionality of the PCL-R and 11 studies assessing the factor structure of the SRP-III and SRP-SF were identified.

**Results:** A critical review of the studies revealed inconsistent findings as to the underlying structure of the PCL-R and SRP-III/SF. Research has been limited by methodological and conceptual weaknesses, which calls into question the applicability of its findings. As such, it is suggested that prior results should be interpreted with caution.

**Conclusion:** Future research should test competing models derived on the basis of previous research and theory, report the results of a differential predictive validity or alternative test, provide all relevant fit indices, utilize new data sets of appropriate size, avoid parceling procedures with short scales, and report the results of composite reliability.

**Keywords:** psychopathy; construct validity; PCL-R; SRP-III/SF; differential predictive validity
Introduction

The concept of psychopathy has been difficult to operationalize and research in the area of psychopathy measurement is compromised by the absence of an agreed definition of the disorder (O’Kane, Fawcett, & Blackburn, 1996). The description of psychopathy which has received the most widespread acceptance among researchers and clinicians is the one proposed by Cleckley (1941). Cleckley suggested psychopathy to be composed of 16 traits reflecting affective and interpersonal deficits, including callousness, lack of guilt, and egocentricity (see da Silva, Rijo, & Salekin, 2012 for a detailed description). This characterization of psychopathy has served as the basis for creating the Psychopathy Checklist (PCL; Hare, 1980) and its updated version, the Psychopathy Checklist – Revised (PCL-R; Hare, 1991, 2003) – often referred to as the “gold standard” for measuring psychopathy in clinical and forensic settings.

The PCL-R

The PCL-R (Hare, 1991) is a 20-item scale completed by a trained administrator on the basis of interviews and case-history data. All items are rated on a 3-point scale (0 = does not apply, 1 = applies to a certain extent, 2 = definitely applies), with scores varying from 0 to 40. A cut-off score of 30 has been suggested for diagnosing psychopathy (Hare & Neumann, 2008). Although the measure was first developed and validated using data from North American samples of male offenders and forensic psychiatric patients, more recent research reported the instrument’s reliability and validity among offender samples from other cultural backgrounds (e.g., Grann, Långström, Tengström, & Kullgren, 1999), adolescent offenders (e.g., Forth & Burke, 1998; Forth & Mailloux, 2000), female offenders (e.g., Salekin, Rogers, & Sewell,
It has been noted that psychopathy, as indexed by the PCL-R and its progeny, can predict violent recidivism (see Dhingra & Boduszek, 2013 for a review; Hart, Kropp, & Hare, 1988; McCuish, Corrado, Hart, & DeLisi, 2015; Serin, 1996; Serin & Amos, 1995; Serin, Peters, & Barbaree, 1990) and sexual reoffending (Furr, 1993; Olver & Wong, 2015; Quinsey, Rice, & Harris, 1995; Rice, Harris, & Quinsey, 1990), which urged Rice and Harris (1995) to propose that the instrument should be used in clinical and legal decision-making. However, a growing body of evidence suggests that this predictive utility for crime is largely attributable to factor 2 (lifestyle/antisocial) rather than factor 1 (affective/interpersonal) scores (Salekin, Rogers, & Sewell, 1996; Skeem & Mulvey, 2001).

The PCL-R ratings were proposed to be best captured by two-, three-, and four-factor models. The two-factor solution is composed of two distinct yet correlated facets, namely factor 1 (affective/interpersonal) and factor 2 (lifestyle/antisocial) (Harpur, Hakstian, & Hare, 1988; Harpur, Hare, & Hakstian, 1989). This solution, however, was not replicated in studies among female (e.g., Salekin et al., 1997) and African-American offenders (Kosson, Smith, & Newman, 1990). The four-factor conceptualization of psychopathy is underpinned by interpersonal, affective, lifestyle, and antisocial facets (Hare, 2003; Hare & Neumann, 2006). Based on 13 PCL-R items, Cooke and Michie (2001) argued for a three-factor hierarchical model, incorporating interpersonal (deceitful interpersonal style), affective (deficient affective experience), and behavioral (impulsive and irresponsible behavioral style) dimensions. This three-factor solution omits items referring to criminal/antisocial behavior, which may be a strong correlate of psychopathy rather than its integral part (Skeem & Cooke, 2010a, b). Finally, more recent factor analytic work revealed the
superiority of bifactor models in grasping the instrument’s dimensionality (Flores-Mendoza, Alvarenga, Herrera, & Abad, 2008; Patrick, Hicks, Nichol, & Krueger, 2007).

Despite the well-documented predictive utility of the PCL-R, its construct validity remains debatable. Although the intention was for the PCL-R items to reflect Cleckley’s original conceptualization of psychopathic personality, the formulation of psychopathy as grasped by the measure appears to be weighted more heavily towards indicators of behavioural expressions of the disorder, such as deviancy and maladjustment (Edens, Skeem, Cruise, & Cauffman, 2001; Patrick, 2007; Patrick et al., 2007; Rogers, 1995). Harpur, Hare, and Hastian (1989) argued that the PCL factor 1 corresponds with the classic depiction of psychopathy, whereas factor 2 is more closely related with the measures of criminal behaviour and Antisocial Personality Disorder (APD). Empirical research demonstrated that only factor 1 items function equivalently well across race and gender (e.g., Bolt, Hare, Vitale, & Newman, 2004; Cooke, Kosson, & Michie, 2001); poor generalizability of factor 2 was reported for substance-dependent patients (McDermott et al., 2000). Notably, antisocial traits were found to decline over time (Blonigen, Hicks, Krueger, Patrick, & Iacono, 2006; Gill & Crino, 2012). These findings may suggest that affective/interpersonal items lie closer to the core of psychopathy.

The SRP-III and SRP-SF

Notwithstanding the value of clinician-administered measures, their use is time-consuming and requires extensive training. Further, although detailed clinical history can be obtained for participants recruited in clinical settings, such information does not usually exist or is accessible for subclinical samples (Lilienfeld & Fowler, 2007).
With these limitations in mind, Hare and colleagues created a self-report version of the PCL(-R), the Self-Report Psychopathy Scale (SRP). The first edition of the SRP (Hare, 1985) consisted of 29 items, however, it failed to adequately address the core features of a psychopathic personality, such as callousness and dishonesty (Lilienfeld & Fowler, 2007). The SRP-II was composed of 60 items, 31 of which formed the core of the scale and aligned with the two factors of the PCL-R (Williams & Paulhus, 2004). Hare (2003), in a validation study within a forensic sample, reported a moderate correlation between the SRP-II and PCL-R ($r = .54$). The latest version of the measure, the SRP-III (Paulhus, Neumann, & Hare, in press) consists of 64 items measured on a five-point Likert scale. Paulhus et al. also developed a shortened, 29-item form of the scale (SRP-SF) in order to reduce the administration time.

The SRP-III factor scores were positively correlated with verbal and physical bullying, drug use, thrill seeking, and aggression (Debowska, Boduszek, Kola, & Hyland, 2014; Gordts, Uzieblo, Neumann, Van den Bussche, & Rossi, in press; Neal & Sellbom, 2012), and negatively correlated with empathy, honesty, conscientiousness, and agreeableness (Neal & Sellbom, 2012; Seibert, Miller, Few, Zeichner, & Lynam, 2011). As for the underlying factor structure, both the SRP-III and SRP-SF were best captured by four correlated facets, including interpersonal manipulation (IPM), callous affect (CA), erratic lifestyle (ELS), and antisocial behavior (ASB). However, recent analytic work has revealed some inconsistent results as to the dimensionality of the SRP-III (e.g., Debowska et al., 2014; Seibert et al., 2011), and unsatisfactory model fit parameters when using scale items as indicators (Debowska et al., 2014; Neal & Sellbom, 2012).

Importantly, it has been stipulated that increased psychopathic traits, such as dominance and manipulativeness, can be found among individuals representing non-
criminal settings where impersonal style and cold calculation are valued (e.g., business, law enforcement, and politics) (Hall & Benning, 2007; Lilienfeld et al., 2012), indicating that criminal/antisocial behavior does not constitute an essential part of the construct of psychopathy. Consequently, because the SRP-III and SRP-SF were generated on the basis of the PCL-R and hence contain items referring to criminal/antisocial conduct, their suitability for use with non-forensic populations appears limited.

The current study

As noted above, psychopathy is presented as a complex set of dimensions, which renders the disorder difficult to capture and define (Ogloff, 2006). This difficulty is further compounded by the lack of agreement among researchers as to what constitutes the essence of psychopathy. Indeed, some researchers have argued that criminal/antisocial behaviors form a critical/important part of the disorder (e.g., Hare & Neumann, 2005; Neumann, Hare, & Pardini, 2014; Vitacco, Neumann, & Jackson, 2005a); others have suggested that criminal/antisocial tendencies are the outcome of psychopathic traits (e.g., Boduszek, Dhingra, Hyland, & Debowska, 2015; Cooke & Michie, 2001; Skeem & Cooke, 2010a, b). The conceptual quandary must be resolved to enable the development of reliable and valid tools for the assessment of psychopathy. Despite the conceptual confusion surrounding the PCL-R and its self-report analogue, the SRP-III, they remain the most widely used measures of psychopathy in both research and clinical practice (Dhingra & Boduszek, 2013; Lee & Ashton, 2005). Additionally, the PCL-R is often equated with the concept of psychopathy, which is evidenced by its use as a referent for estimating the construct validity of other measures of the disorder (e.g., Brinkley, Schmitt, Smith, & Newman, 2001; Poythress et al., 2010). In light of the great theoretical importance assigned to
the PCL-R, there exists a need for further evaluation of the measure and its derivatives. Although research revealed inconsistent results as to the dimensionality, reliability, and differential predictive validity of the PCL-R and SRP-III, a detailed critical appraisal of this prior work is missing. Here, we aim to address this gap by reviewing, summarizing, and critically engaging with the most recent findings into the dimensionality of the PCL-R, SRP-III, and SRP-SF. Based on our analysis results, we provide a set of directions for surpassing the current methodological and conceptual limitations in the field of psychopathy measurement.

Methodology

Search strategy

A search in PubMed, PsychInfo, Scopus, Web of Science, and Science Direct was performed in August, 2015. The following keywords were used in order to identify relevant articles: psychopathy, Psychopathy Checklist-Revised, PCL-R, Self-Report Psychopathy Scale, SRP-III, SRP-IV, Self-Report Psychopathy Scale-Short Form, SRP-SF, combined with factor structure, factor analysis, dimensionality, reliability, and (construct) validity. Google Scholar and Robert Hare's website devoted to the study of psychopathy (www.hare.org) were searched for complementary literature to ascertain that all relevant materials were found. Cited published research not generated in the search was also accessed.

Selection process

Articles reviewed in the current study met the following selection criteria:
1. The study assessed factor structure of at least one of the following psychopathy scales: PCL-R, SRP-III, and SRP-SF.

2. The study is an original piece of research with primary or secondary data analysis (meta-analyses were excluded).

3. The study was written in English and published in a peer-reviewed journal.

4. Given a plethora of studies examining construct validity and dimensionality of the PCL-R, only papers published during the last 15 years (2000-2015) were evaluated.

5. Studies assessing the factor structure of the PCL-SV and PCL-YV were excluded for the sake of the brevity of the report.

The abstracts of 324 studies were inspected by both authors in order to ascertain whether they contained relevant information and met all the inclusion criteria. Most studies were rejected because they were not testing construct validity of the measures. Next, the remaining selection criteria were applied to find relevant studies. This process led to the identification of 34 research papers. The methodological quality of the studies was assessed by two independent reviewers. A consensus method was used (i.e., the decision to exclude papers employing item response theory) to resolve disagreements regarding inclusion of a study. Finally, 21 relevant empirical studies examining the construct validity and dimensionality of the PCL-R and 11 studies looking at the factor structure of the SRP-III and SRP-SF were identified.

**Data extraction and analysis**

Relevant information was extracted into summary tables (Tables 1 - 3). The following data from the studies were retrieved: author(s) and year of publication, number of
models tested, best factorial solution, correlations between latent factors, differential predictive validity, sample, and method of testing. Findings are presented in separate tables for each of the reviewed measures. Additionally, a narrative review of the results is provided.

**Results**

**PCL-R**

Results of 21 studies examining factor structure of the PCL-R are presented in Table 1 below. As shown in the table, research findings are inconsistent as to the underlying structure of the PCL-R, opening it to further scrutiny. In addition, prior research has been limited by methodological and conceptual weaknesses, which calls into question the applicability of its findings. As such, it is suggested that prior results should be interpreted with caution.

Seven of the reviewed studies examined only one possible model of the PCL-R (Cooke, Kosson, & Michie, 2001; León-Mayer, Folino, Neumann, & Hare, 2015; Medina, Valdés-Sosa, García, Almeyda, & Couso, 2013; Mokros *et al.*, 2011; Neumann, Hare, Johansson, 2013; Neumann *et al.*, 2014; Zwets, Hornsveld, Neumann, Muris, & van Marle, 2015); whereas Neumann, Hare, and Newman (2007) assessed two models, but fit statistics were only reported for one of them. Mokros *et al.* (2011) and Zwets *et al.* (2015) tested the four-factorial solution of the measure. Although the incremental index (CFI) was under the acceptable range (.89) for one of the subsamples in each study, no alternative models were assessed. Moreover, the Tucker-Lewis Index (TLI), another incremental fit index, was not reported (these
particular analyses were conducted using Mplus software and hence the TLI must have been available). Similar problems were demonstrated in Neumann et al.'s (2007) study. Namely, the TLI values for all samples were below the recommended cut-off point of .95 (Hu & Bentler, 1999), whereas the TLI for the forensic psychiatric patients sample fell below the less-conservative acceptable range of .90 (Hoyle, 1995). Cooke et al. (2001), who tested a hierarchical model with one super-ordinate and three subordinate factors, reported the CFI of .92 for both of the samples employed. Neumann et al. (2014), on the other hand, reported the relative fit (one of which was .88) and absolute fit values, but it was not specified exactly which fit indices were used for a particular subsample. It appears, therefore, that all the aforementioned studies sought to confirm the proposed model, rather than validate the measure. Thorough model comparisons as opposed to single model testing are warranted to evaluate the applicability of the above findings.

Zwets et al.'s (2015) approach to modeling the data appears even more surprising in light of the reported correlations between PCL-R facets and external variables (different measures of aggression). Specifically, factor 3 (lifestyle) and factor 4 (antisocial) formed similar significant correlations with external measures except for two instances (those correlations were however in the same direction). This is very problematic as the two dimensions were also very highly correlated with each other ($r = .80$), indicating that a three-factor model should have been considered. Factor 1 (interpersonal) and factor 2 (affective), on the other hand, correlated in different directions with some of the external measures, however, those associations were not statistically significant. This suggests that none of the PCL-R facets had differential predictive validity over the remaining ones, which calls into question the factorial solution tested in the above research. Along similar lines, association
between factor 2 (deficient affective) and factor 3 (impulsive and irresponsible behavior) in Weizmann-Henelius et al.’s (2010) study was very strong (.95), but only factor 3 associated significantly with borderline personality disorder (BPD) (both correlations were positive and the difference in the effect size was small). The aforementioned factors also formed significant positive correlations with APD and cluster B disorders, which indicates very low differential predictive validity.

Neumann et al. (2013), who also reported very high correlations between factor 1 and 2 (.87) as well as factor 3 and 4 (.88), utilized structural equation modelling with low anxiety and fearlessness (LAF) as a criterion variable. It was found that only factor 1 (interpersonal) and factor 4 (antisocial) were significantly positively associated with LAF, indicating that the correlated dimensions measure different concepts. However, these results should be tempered by the fact that the beta values for associations between factor 2 (affective) and 3 (lifestyle) psychopathy and LAF were not reported. Interestingly, when structural equation models were tested for individual PCL-R facets separately, each solution evidenced a good model fit, as indexed by the TLI and RMSEA; however, beta values for those relationships were not reported, therefore, the differential predictive validity of the PCL-R components cannot be established here. Correlations between PCL-R factors were also high in León-Mayer et al. (2015), Mokros et al. (2011) as well as Vitacco, Rogers, Neumann, Harrison, and Vincent’s (2005b) study, yet differential predictive validity or equivalent tests were not performed.

In another study, Medina et al. (2013) tested only the two-factor model of the PCL-R. However, no fit indices were provided and hence it is impossible to comment on the fitness of the solution in this particular study. Medina and colleagues did however report some theoretically unexpected correlations between the PCL-R facets
and external measures. Specifically, both factors and the PCL-R total score formed significant negative correlations with physical aggression, verbal aggression, anger, and hostility. The total PCL-R and its factors correlated in the same direction with all external measures and hence no differential predictive validity was observed.

Additionally, although Cooke, Michie, Hart, and Clark (2005a, b) tested more than one possible solution (two-factor, four-factor [in Cooke et al., 2005b only], and three-factor hierarchical model), fit statistics were not reported for one of the subsamples (North American male offenders and psychiatric patients – a sample used in both studies). Overall, it appears that testing only one possible solution of the PCL-R in the aforementioned studies was misguided and such an approach to data modeling should be avoided in future research, especially if similar problems are encountered. Such uninquisitive examinations of the PCL-R dimensionality may seriously obscure our understanding of the scale’s true factor structure.

Another limitation of some of the reviewed studies pertains to correlating errors of measurement (Hildebrand, Ruiter, Vogl, & Wolf, 2002) and the use of the parceling procedure/testlets\(^1\) (e.g., Cooke & Michie, 2001; Cooke et al., 2005a; Cooke, Michie, & Skeem, 2007; Vitacco et al., 2005b; Weizmann-Henelius et al., 2010). The latter strategy indicates the failure to identify an adequate factorial solution when using individual items of the scale. Parceling technique, developed by Cattell and Burdsal (1975), is sometimes used for scales with multiple indicators in order to reduce the large indicator-to-factor ratio (e.g., Debowska, Boduszek, Kola, & Hyland, 2014; Neal & Sellbom, 2012). However, given that the PCL-R consists of

\(^1\) The terms “parcels” and “testlets” are normally used interchangeably (e.g., Reeve & Lam, 2005). Some researchers, however, refer to parcels as aggregate scores, created prior to model fitting, composed of items from the same content area; whereas testlets are groups of items with local dependence and occur when items associations are too high to be explained by their relationship with a latent trait (Cooke et al., 2007).
only 20 items, it seems that this procedure should not have been used. Since parceling can lead to the acceptance of misspecified models (Kim & Hagtvet, 2003), the use of this technique should be avoided and alternative model solutions ought to be tested if similar problems are encountered in future studies. Weaver, Meyer, Van Nort, and Tristan (2006) applied testlets to two-, three-, four-factor, and two-factor/four-facet models, and concluded that the procedure significantly improved the performance of all solutions, which is not a surprising result.

Flores-Mendoza et al. (2008) and Patrick et al. (2007) utilized bifactorial modeling procedure, which provides an empirically and conceptually distinct alternative to traditional CFA model solutions. Bifactor modeling views covariation among observable indicators to be explained by both “general factors” and “grouping/specific factors” which exist at the same conceptual level. Flores-Mendoza et al. (2008) found a bifactor solution with one general factor and two grouping factors to best capture the data. However, in this particular study, the general and grouping factors had similar factor loadings and the researchers failed to explain the influence of this finding on the subsequent use of the PCL-R in applied settings. Moreover, three of the scale items evidenced non-significant loadings and were removed from the model. Such an approach to data modeling is highly problematic because item removal, especially in established measures, should be guided by the theory. Finally, although the researchers correlated psychopathy dimensions with external variables, it is not clear whether the general factor or the total PCL-R score containing all items was subject to those analyses.

In the study by Patrick et al. (2007), on the other hand, a bifactor model with one general and three grouping factors evidenced the best fit of the data. The general factor and three specific factors showed some differential correlations with
personality trait constructs and externalizing behaviors. The specific affective factor failed to exhibit a single significant association with the measures of normal range personality traits, which may indicate that affect is captured differently by those measures. Although the above studies suggest the utility of applying bifactor modeling, the results are difficult to interpret based on existing theoretical conceptualizations of psychopathy. Specifically, psychopathy has never been theorized to reflect a single latent construct as reflected in models of Patrick et al. (2007) and Flores-Mendoza et al. (2008). Although this solution appears to be based on Cooke and Michie’s (2001) hierarchical model with one super-ordinate and three subordinate facets, the hierarchical model was developed as a combination of theoretical conceptualizations of the nature of psychopathy and statistical procedures that explain the structure of the PCL-R ratings (as suggested by Byrne, 1994). A similar misguided view of the unidimensionality of the psychopathy construct was also assumed by McDermott et al. (2000). Indeed, Bishopp and Hare (2008) suggested that, in his original description of psychopathy, Cleckley was referring to “a constellation or syndrome of personality dimensions, rather than a single defining characteristic” (p. 119).

Another example of the failure to abide by the theoretical conceptualizations of psychopathy is the study by Hildebrand et al. (2002). First, the researchers allowed one of the scale items (impulsivity) to load onto two factors. Second, using exploratory principal components analysis (PCA), an alternative two-factor model was suggested as the best fit for the data. Possible justifications for such a model, however, were not provided and the new factors were not labelled.

Eleven of the reviewed studies tested the three-factor model of psychopathy, where items referring to criminal/antisocial behavior were removed (Cooke et al.,
Cooke & Michie, 2001; Cooke et al., 2005a, b; Cooke et al., 2007; Flores-Mendoza et al., 2008; Hildebrand et al., 2002; Johansson, Andershed, Kerr, & Levander, 2002; Vitacco et al., 2005b; Weaver et al., 2006; Weizmann-Henelius et al., 2010). Cooke et al. (2001, 2005b), and Johansson et al. (2002) found the three-factor or hierarchical three-factor model with 13 items to be the best model fit for the data, whereas the same models with testlets were reported as the best factorial solutions in six other studies (Cooke & Michie, 2001; Cooke et al., 2005a; Cooke et al., 2007; Vitacco et al., 2005b; Weaver et al., 2006; Weizmann-Henelius et al., 2010). Notably, Flores-Mendoza et al. (2008), who assessed both the three-factor and bifactorial model, argued for the superiority of the latter. Further, although Patrick et al. (2007) did not examine a three-factor model, they established a bifactor solution with one general and three specific factors as the best model fit. This indicates the need to further explore the possibility that psychopathy is best captured by general and grouping factors existing at the same conceptual level.

An alternative approach to assessing the structure of the PCL-R using multidimensional scaling (MDS) was adopted by Bishopp and Hare (2008). MDS produces a visual output where variables are presented as points in space; from the emergent scalogram, the interpretation of item clusters depends on the theory and the researchers’ judgement (Guttman & Greenbaum, 1998). Based on such a subjective interpretation, Bishopp and Hare (2008) suggested that the PCL-R is best captured by four distinct factors (as evidenced by the three-dimensional MDS solution). The two-dimensional MDS solution, however, could be interpreted with respect to the two- and four-factor models suggested within literature. “Factors 1 and 2 of the two-factor model are marked along the horizontal axis while the four-factor model is indicated vertically as F1, F2, F3 and F4” (p. 124). It may be, thus, that both two and four
factors of psychopathy exist simultaneously at equal conceptual footing and compete for explaining item variance – a solution referred to as the multitrait-multiplemethod (MTMM) model, which can be tested using CFA and SEM techniques (Maas, Lensvelt-Mulders, & Hox, 2009). Worthy of note, MTMM models have been previously found to best represent the dimensionality of two measures derived from the PCL-R, namely the PCL-SV (Boduszek et al., 2015) and the SRP-III (Debowska et al., 2014). Similar analyses using the PCL-R are lacking but, it appears, should be conducted.

The final issue pertains to the samples used in the reviewed studies. Specifically, a number of studies were conducted using secondary data and some of the samples were used repeatedly for the purpose of examining the same factorial solutions. Three of the reviewed studies were conducted using North American data set of male offenders described by Hare (2003) (Bishopp & Hare, 2008; Neumann et al., 2007; Neumann et al., 2014). Three quarters of Cooke and Michie’s (2001) participants included in study 1, 2, and 3 were derived from Hare’s (1991) PCL-R standardization sample. The same data from Scottish male prisoner samples were utilized in three of the reviewed studies (Cooke & Michie, 2001, study 4; Cooke et al., 2005a; Cooke et al., 2007). The sample of North American male adult offenders and psychiatric patients \((N = 2,067)\), obtained for the purpose of prior research, were employed by both Cooke et al. (2005a) and Cooke et al. (2005b). It appears, therefore, that the results of the above studies should be taken with caution and more research using new data sets is needed to verify those findings.

---

2 In both of those studies, the MTMM model was referred to as a bifactor model with two general and four grouping factors.
## Table 1

*Studies assessing factor structure of the PCL-R*

<table>
<thead>
<tr>
<th>Author(s) and year of publication</th>
<th>No. of models tested</th>
<th>Best factorial solution</th>
<th>Correlations between factors</th>
<th>Differential predictive validity</th>
<th>Sample</th>
<th>Method of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bishopp &amp; Hare (2008)</td>
<td>n/a</td>
<td>Four factors</td>
<td>n/a</td>
<td>n/a</td>
<td>4,630 male offenders</td>
<td>MDS</td>
</tr>
<tr>
<td>Cooke et al. (2001)</td>
<td>1</td>
<td>Hierarchical model with 3 subordinate factors and 1 super-ordinate factor (13 items)</td>
<td>n/a</td>
<td>n/a</td>
<td>359 Caucasian and 356 African American inmates</td>
<td>CFA, IRT</td>
</tr>
<tr>
<td>Cooke &amp; Michie (2001)</td>
<td>Study 1 = 2 Study 2 = 5 Study 3 = 1 Study 4 = 1</td>
<td>STUDY 1: none but 2-factor model better than 1-factor model STUDY 2: 3 factors with testlets (13 items) – superordinate factor was included STUDY 3: 3 factors with testlets (13 items) – hierarchical model with 1 superordinate factor STUDY 4: 3 factors with testlets (13 items) – hierarchical model with 1 superordinate factor</td>
<td>n/a</td>
<td>n/a</td>
<td>STUDY 1, 2, &amp; 3: 2,067 North American inmates and forensic psychiatric patients STUDY 4: 596 male Scottish prisoners</td>
<td>EFA, CFA, IRT</td>
</tr>
<tr>
<td>Cooke et al. (2005a)</td>
<td>2</td>
<td>3-factor hierarchical model with testlets</td>
<td>n/a</td>
<td>n/a</td>
<td>1,316 British adult male offenders (some secondary data); 2,067 North American adult male offenders</td>
<td>CFA, IRT</td>
</tr>
<tr>
<td>Cooke et al. (2005b)</td>
<td>3</td>
<td>3-factor hierarchical model with testlets</td>
<td>n/a</td>
<td>n/a</td>
<td>2,067 North American adult male offenders and psychiatric patients; 1,563 adult male offenders and forensic psychiatric patients</td>
<td>CFA, IRT</td>
</tr>
<tr>
<td>Study</td>
<td>Factors</td>
<td>Model Description</td>
<td>Parameters</td>
<td>Sample Details</td>
<td>Methodology</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Cooke et al. (2007)</td>
<td>11</td>
<td>Hierarchical 3-factor with testlets</td>
<td>n/a</td>
<td>1,212 adult male offenders from continental Europe</td>
<td>CFA</td>
<td></td>
</tr>
<tr>
<td>Flores-Mendoza et al. (2008)</td>
<td>7</td>
<td>Bi-factor model with 2 specific factors and 1 general factor (17 items)</td>
<td>F1 &amp; F2 (.37)</td>
<td>Weak differential predictive validity 124 Brazilian male prisoners</td>
<td>CFA</td>
<td></td>
</tr>
<tr>
<td>Hildebrand et al. (2002)</td>
<td>7</td>
<td>Alternative 2-factor model</td>
<td>F1 &amp; F2 (.25)</td>
<td>n/a 98 male forensic psychiatric patients</td>
<td>CFA (ML) and exploratory PCA using oblique rotation</td>
<td></td>
</tr>
<tr>
<td>Johansson et al. (2002)</td>
<td>2</td>
<td>3-factor model (13 items)</td>
<td>n/a</td>
<td>293 adult male violent offenders from a maximum-security prison in Sweden</td>
<td>EFA, CFA (WLS)</td>
<td></td>
</tr>
<tr>
<td>León-Mayer et al. (2015)</td>
<td>1</td>
<td>4-factor model</td>
<td>Range = .42 -.89</td>
<td>n/a 209 male inmates from the Prison of Los Andes, Chile</td>
<td>CFA (WLSMV)</td>
<td></td>
</tr>
<tr>
<td>McDermott et al. (2000)</td>
<td>n/a</td>
<td>2-factor model (prisoner sample); 1-factor model (substance-dependent sample)</td>
<td>n/a</td>
<td>326 male inmates from a prison in southern Wisconsin; 620 (n = 442 men, n = 178 women) substance-dependent patients</td>
<td>Exploratory orthogonal and oblique common factoring</td>
<td></td>
</tr>
<tr>
<td>Medina et al. (2013)</td>
<td>1</td>
<td>2-factor model (2 items did not load on any of the factors)</td>
<td>n/a</td>
<td>Lack of differential predictive validity 124 Cuban violent offenders</td>
<td>Factor analysis with normalized varimax rotation</td>
<td></td>
</tr>
<tr>
<td>Mokros et al. (2011)</td>
<td>1</td>
<td>4-factor model (18 items)</td>
<td>Range for North American sample = .59 -.86 Range for German sample = .83 -.94</td>
<td>n/a North American sample: 2,622 male offenders (secondary data) German sample: 443 male offenders</td>
<td>CFA</td>
<td></td>
</tr>
<tr>
<td>Neumann et al.</td>
<td>1</td>
<td>4-factor model (18 items)</td>
<td>Range = .59 -.88</td>
<td>Lack of differential predictive validity 1,031 offenders who underwent</td>
<td>CFA</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>N (participants)</td>
<td>Model/Methodology</td>
<td>Validity/Assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neumann et al. (2007)</td>
<td>4,865 male offenders (sample 1), 1,099 female offenders (sample 2), 965 forensic psychiatric patients (sample 3)</td>
<td>4-factor model</td>
<td>CFA (MLR)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neumann et al. (2014)</td>
<td>(1) 4,685 North American adult male offenders; (2) 209 South American adult male offenders; (3) 1,983 European male offenders; (4) 3,389 adult males from a psychiatric hospital in the Netherlands; (5) 1,046 Austrian male offenders; (6) 1,031 Swedish male offenders; (7) 445 Finnish male offenders; (8) 965 male psychiatric patients; (9) 1,099 female offenders; (10) 203 adults from corporate settings; (11) 304 male offenders from prisons in Wisconsin; (12) 208 young adult males from the Pittsburgh Youth Study; (13) 12,201 participants mega-sample: samples 1, 3, &amp; 4 combined</td>
<td>4-factor model</td>
<td>SEM (WLSMV)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patrick et al. (2007)</td>
<td>593 male inmates recruited from 2 American prisons</td>
<td>Bifactor with 1 general factor and 3 specific factors (20 items)</td>
<td>EFA CFA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitacco et al. (2005b)</td>
<td>96 male mentally disordered offenders from the Tarrant County Jail in Fort Worth, Texas</td>
<td>3-factor testlets model</td>
<td>CFA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weaver et al. (2006)</td>
<td>1,566 male sex offenders</td>
<td>3-factor testlets model (13 items, 6 testlets)</td>
<td>CFA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weizmann-Henelius et al. (2007b)</td>
<td>97 female homicide offenders (data collected from archives; file-only)</td>
<td>3-factor model with 6 testlets (13 items)</td>
<td>CFA (ML)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>al. (2010)</td>
<td>1</td>
<td>4 factors</td>
<td>Range = .37 - .84</td>
<td>Moderate differential predictive validity</td>
<td>411 forensic psychiatric inpatients – personality disorders group (n = 269), psychiatric disorders group (n = 142)</td>
<td>CFA</td>
</tr>
</tbody>
</table>

**Note.** CFA = Confirmatory factor analysis; EFA = Exploratory factor analysis; MDS = Multidimensional scaling; ML = Maximum likelihood; PCA = Principal components analysis; SEM = Structural Equation Modelling; WLS = Weighted least square estimation; WLSMV = Robust weighted least squares estimation.
SRP-III and SRP-SF

Results of 11 studies evaluating the dimensionality and construct validity of the SRP-III and SRP-SF are presented in Tables 2 and 3 respectively.

Two of the reviewed studies (Mahmut, Menictas, Stevenson, & Homewood, 2011; Neal & Sellbom, 2012) revealed the four-factor model of the SRP-III, composed of interpersonal manipulation (IPM), callous affect (CA), erratic lifestyle (ELS), and antisocial behavior (ASB), to be the best fit for the data. In five other studies (Declercq, Carter, & Neumann, 2015; Gordts et al., in press; León-Mayer et al., 2015; Neumann et al., 2014; Neumann, Schmitt, Carter, Embley, & Hare, 2012), the SRP-III and SRP-SF demonstrated a good fit for the four-factor model, however, no competing model solutions were assessed. The three-factor model suggested for the PCL-R scores by Cooke and Michie (2001) was found inappropriate for the SRP-III (e.g., Mahmut et al., 2011), indicating that criminal/antisocial tendencies may constitute an integral part of psychopathy construct as indexed by this self-report instrument.

Debowska et al. (2014), Gordts et al. (in press) as well as Neal and Sellbom (2012) failed to find an appropriate model fit when using all 64 SRP-III items as indicators. Similarly to some prior studies examining the factor structure of the PCL-R, the researchers employed the parceling technique in order to reduce model complexity. The 64 items were assigned randomly into 16 radical parcels (four for each factor), which increased the CFI and TLI values and decreased the RMSEA and SRMR values for the assessed solutions. Gordts et al. (in press) and Neal and Sellbom (2012) suggested the four-factor parcelled model as the best fit for the data, with the CFI value of .92 and .95 respectively (the TLI values were not reported). Debowska et
al. (2014), who used a Polish version of the SRP-III, assessed a bifactor conceptualization of psychopathy with two general factors (interpersonal/affective and lifestyle/antisocial) and four grouping factors (interpersonal manipulation, callous affect, erratic lifestyle, and antisocial behavior) and found it to be a statistically superior representation of the data (CFI = .96, TLI = .94, RMSEA = .058, SRMR = .045). Standardized factor loadings were significantly stronger for the grouping factors than for the general factors, therefore, the Polish SRP-III was suggested to be measuring four primary factors of psychopathy and two hidden factors. Given that all studies examining factor structure of the SRP-III and PCL-R reviewed here which have tested a bifactor model revealed it to be the best fit for the data (Debowska et al., 2014; Flores-Mendoza et al., 2008; Patrick et al., 2007), future research should include a bifactorial conceptualization as a comparison model.

Further, despite some high correlations between the four SRP-III factors, they evidenced good differential predictive validity (e.g., Debowska et al., 2014; Neal & Sellbom, 2011), suggesting that they measure disparate dimensions. However, in León-Mayer et al.’s (2015) investigation into the SRP-SF dimensionality, correlations between the IPM and CA factors and between the IPM and ELS factors were as high as .99 and .93 respectively, which questions their conceptual distinctiveness. Declercq et al. (2015) argued that strong correlations between the SRP-SF dimensions provide evidence for the existence of a super-ordinate psychopathy factor, however, neither a hierarchical nor a bifactor solution was tested to support this proposition. Interestingly, Gordts et al. (in press) reported corresponding SRP-III and SRP-SF facets to form some different associations with external variables. For example, SRP-III-IPM, but not SRP-SF-IPM, correlated significantly with the Adolescent Peer Relations Instrument – Social Target subscale (Parada, 2000). The SRP-III-CA
associated positively and SRP-SF-CA associated negatively with the Adolescent Peer Relations Instrument – Verbal Target subscale, yet these correlations were statistically non-significant. These differential correlations however indicate that the two versions of the SRP may not be qualitatively equal.

Mahmut et al. (2011) tested the viability of the SRP-III as a PCL-R-analogous instrument of psychopathy within a community sample (N = 500). Although the original scale consists of 64 items, the researchers dropped 24 items with loadings less than .30 in the exploratory factor analysis (EFA). As explained in the previous section, such an approach to factor structure examination is misguided as the resulting content of the measure is based on statistical rather than theoretical superiority. The EFA technique was also employed in Freeman and Samson’s (2012) study. Psychopathy as indexed by the SRP-III was suggested to be composed of four factors, but some cross-loadings between the IPM and CA facets were evident. The utilization of the EFA technique in both of the above studies does not seem justified. This is because, unless no preconceived factor structure is provided, exploratory techniques should be avoided when assessing dimensionality of established instruments.

Seibert et al. (2011), on the other hand, employed the EFA to examine the joint factor structure of three self-report measures of psychopathy, namely the SRP-III, the Levenson Self-Report Psychopathy Scale (LSRP; Levenson, Kiehl, & Fitzpatrick, 1995), and the Psychopathic Personality Inventory-Revised (PPI-R; Lilienfeld & Widows, 2005). The analysis was performed on 14 psychopathy subscales and the domains from the five-factor personality model (neuroticism, extraversion, openness, agreeableness, and conscientiousness) together. Results revealed the existence of four psychopathy facets, however, not all SRP-III subscales loaded significantly on the expected factors. Specifically, both IPM and CA subscales
loaded onto factor 1 (.86 and .47 respectively), conceived as representing interpersonal/affective deficits. The CA subscale also loaded negatively on factor 4 (-.55), which represented the lack of emotion. The ELS subscale loaded on factor 3 (.86), which captured elements of poor impulse control. The ASB subscale did not load on any factor, yet this result was not explored further. Given existing theoretical conceptualizations of psychopathy, findings of this particular study are difficult to interpret, however, it appears that antisocial tendencies share little variance with the remaining psychopathy scales.

Neumann and Pardini’s (2014) study provides another example of an analysis used to simultaneously evaluate the latent item measurement structures of two scales, the SRP-SF (ASB items were omitted) and the Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Lavender, 2002). Although the solution with six latent factors evidenced an adequate model fit, full parameter results are only available upon request. Moreover, the omission of items referring to antisocial tendencies, which form an integral part of the measure, does not allow for inferences regarding the SRP-SF reliability and dimensionality to be made.

Indeed, an important limitation identified in studies utilizing the SRP-SF pertains to the number of scale items used. Although the original SRP-SF consists of 29 items, some researchers have reduced the number of indicators. For example, Gordts et al. (in press) used a 28-item scale. Neumann et al. (2014) utilized a 19- and 26-item (but the figure provided suggests that only 18 items were included) SRP-SF, without explaining which scale items were excluded. Neumann et al. (2012) analyzed the factor structure of an experimental 19-item version of the SRP. This lack of consistency significantly undermines the generalizability of research findings and the reliability of the measure.
Table 2

Studies assessing factor structure of the SRP-III

<table>
<thead>
<tr>
<th>Author(s) and year of publication</th>
<th>No. of models tested</th>
<th>Best factorial solution</th>
<th>Correlations between factors</th>
<th>Differential predictive validity</th>
<th>Sample</th>
<th>Method of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debowska et al. (2014)</td>
<td>8</td>
<td>Bi-factor model with 2 general and 4 grouping factors (64 items, 16 parcels)</td>
<td>Range = .57 - .88</td>
<td>Moderate differential predictive validity</td>
<td>319 Polish working adults</td>
<td>CFA (MLR)</td>
</tr>
<tr>
<td>Freeman &amp; Samson (2012)</td>
<td>n/a</td>
<td>4-factor (64 items)</td>
<td>Range = .51 - .74</td>
<td>Lack of differential predictive validity</td>
<td>Australian community sample (N = 300)</td>
<td>EFA</td>
</tr>
<tr>
<td>Gordts et al. (in press)</td>
<td>2</td>
<td>4-factor model (64 items, 16 parcels)</td>
<td>Range = .49 - .74</td>
<td>Moderate differential predictive validity</td>
<td>Belgian community sample (N = 1,510)</td>
<td>CFA (WLSMV for items as indicators &amp; RML for parcels)</td>
</tr>
<tr>
<td>Mahmut et al. (2011)</td>
<td>4</td>
<td>4-factor model (40 items)</td>
<td>Range = .24 - .51</td>
<td>n/a</td>
<td>Community sample (N = 500)</td>
<td>EFA, CFA</td>
</tr>
<tr>
<td>Neal &amp; Sellbom (2012)</td>
<td>8</td>
<td>4-factor model (64 items, 16 parcels)</td>
<td>Range = .62 - .76</td>
<td>Good differential predictive validity</td>
<td>602 college students</td>
<td>CFA</td>
</tr>
<tr>
<td>Seibert et al. (2011)</td>
<td>n/a</td>
<td>4-factor model (64 items)</td>
<td>Range = .33 - .59</td>
<td>Weak differential predictive validity</td>
<td>143 undergraduate psychology students from a large Southeastern university</td>
<td>EFA with an oblimin rotation</td>
</tr>
</tbody>
</table>

Note. CFA = Confirmatory factor analysis; EFA = Exploratory factor analysis; RML = Robust maximum likelihood; SEM = Structural Equation Modelling; WLSMV = Robust weighted least squares estimation.
### Table 3

**Studies assessing factor structure of the SRP-SF**

<table>
<thead>
<tr>
<th>Author(s) and year of publication</th>
<th>Models tested</th>
<th>Best factorial solution</th>
<th>Correlations between factors</th>
<th>Differential predictive validity</th>
<th>Sample</th>
<th>Method of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Declercq et al. (2015)</td>
<td>1</td>
<td>4-factor model (29 items)</td>
<td>Range = .70 - .88</td>
<td>n/a</td>
<td>Dutch-speaking female undergraduates (N = 343)</td>
<td>CFA</td>
</tr>
<tr>
<td>Gordts et al. (in press)</td>
<td>1</td>
<td>4-factor model (28 items)</td>
<td>Range = .64 - .84</td>
<td>Moderate differential predictive validity</td>
<td>Belgian community sample (N = 1,510)</td>
<td>CFA (WLSMV)</td>
</tr>
<tr>
<td>León-Mayer et al. (2015)</td>
<td>1</td>
<td>4-factor model (29 items)</td>
<td>Range = .85 - .99</td>
<td>n/a</td>
<td>209 male inmates from the Prison of Los Andes, Chile</td>
<td>CFA (WLSMV)</td>
</tr>
<tr>
<td>Neumann et al. (2014)</td>
<td>1</td>
<td>4-factor model (number of items included differed across samples)</td>
<td>Range for sample 1 only = .79 - .88</td>
<td>n/a</td>
<td>(1) North American adult mega sample (N = 1,730) – 26-item SRP-SF; (2) World adult mega sample (N = 33,016) – 19-item SRP-SF; (3) Male offenders from prisons in Wisconsin (N = 304) – number of SRP-SF items not specified; (4) Young adult males from the Pittsburgh Youth Study (N = 208) - number of SRP-SF items not specified</td>
<td>SEM (WLSMV)</td>
</tr>
<tr>
<td>Neumann &amp; Pardini (2014)</td>
<td>1</td>
<td>6-factor model (SRP-SF-ASB items were excluded; CFA performed jointly on SRP-SF and YPI items)</td>
<td>n/a</td>
<td>Lack of differential predictive validity</td>
<td>Young adult males from the Pittsburgh Youth Study (N = 425)</td>
<td>CFA (WLSMV)</td>
</tr>
<tr>
<td>Neumann et al. (2012)</td>
<td>1</td>
<td>4-factor model (19 items)</td>
<td>Range for total sample = .40 - .80; Range for female sample = .40 - .77; Range for male sample = .33 - .79</td>
<td>Moderate differential predictive validity</td>
<td>34,118 participants (14,301 men and 19,817 women) from the International Sexuality Description Project-2 (ISDP-2)</td>
<td>SEM (WLSMV)</td>
</tr>
</tbody>
</table>

*Note. ASB = Antisocial Behavior; CFA = Confirmatory factor analysis; CFI = Comparative Fit Index; SEM = Structural Equation Modelling; WLSMV = Robust weighted least squares estimation; YPI = Youth Psychopathic Traits Inventory.*
Discussion

The PCL-R and its self-report analogue, the SRP-III/SF, are the measures of psychopathy most commonly used in research and clinical practice. Although empirical studies have consistently demonstrated that increased PCL-R scores can predict violence and general recidivism (Dhingra & Boduszek, 2013; Salekin et al., 1996), the instrument’s construct validity has recently been challenged (Cooke & Michie, 2001; Skeem & Cooke, 2010a, b). Inconsistencies have also been reported regarding the factor structure of the measures, with some researchers suggesting two- (e.g., McDermott et al., 2000; Medina et al., 2013), three- (e.g., Cooke et al., 2005a, b), four- (e.g., Freeman & Samson, 2012; León-Mayer et al., 2015; Neumann et al., 2013, 2014; Zwets et al., 2015), and bi-factor (e.g., Debowska et al., 2014; Flores-Mendoza et al., 2008; Patrick et al., 2007) models to best capture the PCL-R and SRP-III/SF ratings. In spite of this conflicting evidence, there is a lack of a critical review of such research. It appears that our understanding of the dimensionality of the instruments may be adversely influenced by the application of inappropriate methods for examining their latent structure. Since the PCL-R is often equated with the concept that it contends to measure, such methodological limitations may subsequently affect our understanding of psychopathy. In light of the above, the objective of the present paper was to provide a critical evaluation of the most recent studies assessing the factor structure of the PCL-R and SRP-III/SF. We conclude with recommendations for future scholarship in the field of psychopathy measurement.

As presented in detail in the results section, one of the reasons for contradictory findings amongst psychopathy factor analytic literature may be the failure to compare a number of alternative models. Although the dimensionality of the PCL-R and its derivatives remains debatable, we found a significant number of
studies to assess one possible solution (e.g., Cooke et al., 2001; Declercq et al., 2015; León-Mayer et al., 2015; Neumann et al., 2013, 2014), even when fit indices for the tested model were below the recommended values (such as in studies by Mokros et al., 2011; Zwets et al., 2015). Of questionable value is also research where fit statistics were reported selectively (e.g., Gordts et al., in press; Neumann et al., 2014; Vitacco et al., 2005; Zwets et al., 2015) or not reported at all (e.g., Medina et al., 2013). In addition, it appears that researchers underestimate the usefulness of bifactor modeling in grasping the PCL-R and SRP-III/SF scores, even though bifactorial solutions evidenced the best model fit in all studies that considered them (Debowska et al., 2014; Flores-Mendoza et al., 2008; Patrick et al., 2007). This finding urged us to suggest that future research using the PCL-R and its derivatives should include a bifactorial conceptualization as a comparison model. Another methodological limitation identified amongst the reviewed studies pertains to the repeated use of the same data sets (e.g., Bishopp & Hare, 2008; Cooke et al., 2005a, b; Neumann et al., 2007; Neumann et al., 2014) and data sets of limited size (e.g., Flores-Mendoza et al., 2008; Hildebrand et al., 2002; León-Mayer et al., 2015; Medina et al., 2013; Seibert et al., 2011; Vitacco et al., 2005; Weizmann-Henelius et al., 2010), possibly resulting in overconfidence in the fitness of some models.

Further, in regards to risk assessment and treatment, it is of utmost importance that those measures also evidence good differential predictive validity. However, only six of the reviewed studies examining the factor structure of the PCL-R (Flores-Mendoza et al., 2008; Medina et al., 2013; Neumann et al., 2013; Patrick et al., 2007; Weizmann-Henelius et al., 2010; Zwets et al., 2015) and eight studies using the SRP-

3 Based on theoretical conceptualizations of psychopathy, we suggest that such a model should comprise two general and four specific factors (also referred to as the MTMM model).
III/SF (Debowski et al., 2014; Freeman & Samson, 2012; Gordts et al., in press; Neal & Sellbom, 2012; Neumann & Pardini, 2014; Neumann et al., 2012; Seibert et al., 2011) reported the results of some differential predictive validity (e.g., differential correlations with external variables). Notably, four of those studies demonstrated a lack of differential predictive validity (Medina et al., 2013; Neumann et al., 2013; Freeman & Samson, 2012; Neumann & Pardini, 2014). Although Neumann et al. (2012) reported a moderate differential predictive validity of the four psychopathy facets, the choice of external correlates included in the study was, to say the least, surprising (e.g., body mass index, adolescent fertility rate, gross domestic product per capita). As such, more research into the predictive utility of different psychopathy factors based upon solid theoretical grounds is warranted.

Notwithstanding the importance of factor analytic work in advancing our understanding of psychopathy, it will not reveal the true nature of the construct (Meehl, 1992; Skeem & Cooke, 2010b). In order to avoid exploratory model testing based on statistical procedures only, such studies should adhere to a predefined theoretical framework. Research in the field of psychopathy measurement, however, is compromised by the lack of a clear definition of the disorder (Cooke & Logan, 2015; Corrado, DeLisi, Hart, & McCuish, 2015). Further, there is a continuing debate as to whether or not criminal/antisocial tendencies constitute an integral part of psychopathy construct (see discussion between Hare & Neumann, 2010 and Skeem & Cooke, 2010a, b). According to the early conceptualization of psychopathy (Cleckley, 1941; Karpman, 1948), the essence of the disorder is characterized by affective and interpersonal deficits, whereas the proneness to transgress social and legal norms seems to comprise a possible behavioural manifestation of psychopathy (Skeem & Cooke, 2010a, b). However, psychopathic traits can make one successful in both
criminal and non-criminal endeavours (Millon, Simonsen & Birket-Smith, 1998). For example, Hassall, Boduszek, and Dhingra’s (2015) research findings revealed that business students, relative to psychology students, scored significantly higher on all four psychopathy dimensions and that psychopathy was a significant correlate of academic achievement. The prevalence of psychopathic traits was also higher in a corporate sample than that found in community samples (Babiak, Neumann, & Hare, 2010). In another study, it was demonstrated that increased psychopathy scores among U.S. presidents were associated with a better-rated presidential performance (Lilienfeld et al., 2012). It appears thus that highly intelligent individuals with psychopathic traits may perform exceptionally well in non-criminal professions which value cold calculation and impersonal style. Along similar lines, Seibert et al.’s (2011) study results revealed that antisocial tendencies, as indexed by the SRP-III, share little variance with psychopathy measures not derived on the basis of the PCL-R. Consequently, in spite of Hare and Neumann’s (2006) assertion that the omission of criminal tendencies from the PCL-R is inconsistent with its structural properties, a growing body of evidence suggests that such tendencies should not be regarded as central to the conceptualization of the construct of psychopathy.

Further, the above stated limitation is especially significant in the evaluation of clinician-administered measures. This is because, although different facets of the disorder are assessed separately, cut-off points used to diagnose psychopathy utilized in such instruments rely on the total scale score rather than ratings obtained on separate dimensions. As such, psychopathy is more likely to be diagnosed in forensic samples, which may be one of the reasons why psychopaths are overrepresented in prisons. Indeed, research indicates that approximately 25 per cent of inmates meet diagnostic criteria for psychopathy (Lilienfeld & Arkowitz, 2007), compared with 1
per cent of the general population (Coid, Yang, Ullrich, & Hare, 2009). It appears thus that the inclusion of criminal/antisocial tendencies in psychopathy measures constitutes a serious drawback which ought to be addressed. The use of similar items in self-report measures is perhaps less problematic because such scales are not normally used for diagnostic purposes and hence cut-off points for diagnosis are not specified. Yet, if criminal/antisocial tendencies are just one possible outcome of a psychopathic personality, other non-criminal/antisocial activities in which psychopaths may engage should also be accounted for. It appears, however, that such an approach would be counterproductive and, in terms of research participation, unnecessarily time-consuming.

In consideration of the above, it seems that the PCL-R and its progeny, through concentrating on criminal/antisocial behaviours, do not grasp the essence of psychopathy, as conceptualized by Cleckley (1941). Further, we suggest that, instead of focusing on behaviours associated with psychopathy, be it criminal/antisocial or non-criminal/antisocial, which appear to be potential outcomes of the disorder and not its integral element, psychopathy inventories should only assess relevant psychological traits. It is also of paramount importance that such measures distinguish between Antisocial Personality Disorder (APD; American Psychiatric Association, 2013) and core psychopathy. A clean personality measure of psychopathy uncontaminated with behavioral features would enable researchers to extend the construct to all populations regardless of criminal history (Johansson et al., 2002). Accordingly, notwithstanding the role of the PCL-R and its progeny in forwarding scientific research into psychopathy and its measurement, new generation of research based upon solid theoretical grounds and which “distinguishes between personality deviation and social deviance” is needed (Skeem & Cooke, 2010b, p. 455).
Recommendations for future research

Given the inconsistencies amongst the PCL-R and SRP-III/SF factor analytic literature, there exists a need for providing a set of recommendations which would systemize future research. As such, we recommend that the following indications are used as an absolute minimum when assessing construct validity and dimensionality of psychopathy measures:

1. Confirmatory techniques should be used to test competing models derived on the basis of previous research and theory. It is unacceptable to assume that there is only one solution (e.g., four-factor correlated model). Such an approach to model testing may obscure the true nature of the dimensionality of the measures. Further, bi-factor conceptualization should be used as a comparison model because it helps to assess the validity of a single (or two) general factor, while also acknowledging and incorporating aspects of multidimensionality (grouping factors) (see Hyland, 2015; Reise, Moore, & Haviland, 2010 for more information on the application of bi-factor modeling in psychological research).

2. When the best model fit is multidimensional in nature, a differential predictive validity or alternative test (e.g., correlation with external variables) must be performed in order to verify whether the recognized factors correlate differently with external criteria (the choice of which must be guided by the theory). This issue is especially important when the latent factors are highly correlated (i.e., .50 and above). Very high correlations between factors (such as .70 and above) would indicate the possibility that they reflect the same concept. As suggested by Carmines and Zeller (1979), if highly correlated factors measure different dimensions, they should differently relate to external
variables. If differential predictive validity is not demonstrated, the proposed factorial solution may be yet another “statistical exercise” rather than a proper validation of construct.

3. When assessing the construct validity and dimensionality using the confirmatory factor analysis (CFA), the absolute minimum requirement is that the following fit indices (if available) are provided in order to make direct comparisons between the competing models: the comparative fit index (CFI; Bentler, 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), the root-mean-square error of approximation (RMSEA; Steiger & Lind, 1980), and the standardized root-mean-square residual (SRMR; Bentler, 1995) (all of these are available in Mplus software). If researchers use maximum likelihood (ML) or maximum likelihood robust (MLR) estimators (not recommended with categorical data), they should also report either the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

4. More studies using new data sets of appropriate size should be conducted. As noted by Weaver et al. (2006), much of the research inquiring into the dimensionality of the PCL-R has been based upon samples of questionable size. The requirement of the CFA is that the sample used is of a moderate/large size (Kline, 2010); if this condition is not satisfied, findings may be misleading (Floyd & Widaman, 1995). The minimum sample size required should be calculated based on the number of indicators (test items) and latent factors (Soper, 2015). For example, to test a four-factor model of the PCL-R containing 20 items, the minimum sample size to detect effect is 387 (for the anticipated effect size .10, desired statistical power level .80, and
the probability level .05). For a bi-factor model comprised of two general and
four grouping factors, the sample size required increases to 526.

5. While the use of parceling is acceptable when a scale contains a large number
of items (such as in the case of the 64-item SRP-III), the employment of the
technique with short scales (such as the PCL-R) is questionable and should be
avoided.

6. Finally, composite reliability should be reported instead for internal
consistency (Cronbach’s alpha) in a latent variable modeling context (see
Debowska et al., 2014, p. 235).

References

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

(Eds.), Psychopaths: Current international perspectives (pp. 131–158). The


Bulletin, 217, 238-246.


Birket-Smith, & R.D. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behaviour* (pp. 3-31). New York: Guilford.


Skeem, J. L., & Cooke, D. J. (2010b). One measure does not a construct make: Directions toward reinvigorating psychopathy research—reply to Hare and Neumann (2010). *Psychological Assessment, 22*(2), 455-459.


the development of a four-factor model. *Criminal Justice and Behavior*, 32(5), 526-545.


