Multimethod Assessment of Psychopathy in Relation to Factors of Internalizing and Externalizing From the Personality Assessment Inventory: The Impact of Method Variance and Suppressor Effects

Daniel M. Blonigen
Veterans Affairs Palo Alto Health Care System and Stanford University School of Medicine

Christopher J. Patrick
University of Minnesota, Twin Cities

Kevin S. Douglas
Simon Fraser University

Norman G. Poythress
University of South Florida

Jennifer L. Skeem
University of California–Irvine

Scott O. Lilienfeld
Emory University

John F. Edens
Texas A&M University

Robert F. Krueger
Washington University in St. Louis

Research to date has revealed divergent relations across factors of psychopathy measures with criteria of internalizing (INT; anxiety, depression) and externalizing (EXT; antisocial behavior, substance use). However, failure to account for method variance and suppressor effects has obscured the consistency of these findings across distinct measures of psychopathy. Using a large correctional sample, the current study employed a multimethod approach to psychopathy assessment (self-report, interview and file review) to explore convergent and discriminant relations between factors of psychopathy measures and latent criteria of INT and EXT derived from the Personality Assessment Inventory (Morey, 2007). Consistent with prediction, scores on the affective–interpersonal factor of psychopathy were negatively associated with INT and negligibly related to EXT, whereas scores on the social deviance factor exhibited positive associations (moderate and large, respectively) with both INT and EXT. Notably, associations were highly comparable across the psychopathy measures when accounting for method variance (in the case of EXT) and when assessing for suppressor effects (in the case of INT). Findings are discussed in terms of implications for clinical assessment and evaluation of the validity of interpretations drawn from scores on psychopathy measures.

Keywords: psychopathy, internalizing, externalizing, method variance, suppressor effects

Studies evaluating the validity of psychopathy measures often reveal a theoretically meaningful pattern of convergent and discriminant relations with external criteria across subfactors of these measures (e.g., Hare, 2003; Harpur, Hare, & Hakstian, 1989; Smith & Newman, 1990). The psychopathological domains of internalizing (INT; unipolar depression, anxiety disorders) and externalizing (EXT; antisocial behavior, substance use disorders) illustrate this pattern. Research to date suggests that INT criteria...
evince an inverse relationship with scores on the affective–
interpersonal factor of psychopathy, defined by core personality
features such as superficial charm, grandiosity, manipulation,
and shallow affect (Frick, Lilienfeld, Ellis, Loney, & Silverthorn,
1999; Hicks & Patrick, 2006; Verona, Hicks, & Patrick, 2005), and
a positive association with scores on the social deviance factor,
marked by chronic behavioral features of antisociality (e.g., im-
pulsivity, aggression). Conversely, EXT criteria exhibit modest to
negligible associations with scores on the affective–interpersonal
factor while converging strongly with the social deviance factor
(Hare, 2003; Patrick, Hicks, Krueger, & Lang, 2005). However,
fundamental issues in assessment—such as the method of assess-
ment across predictors and criteria and the moderating role of
suppressor effects on predictor–criterion relations—cloud the con-
sistency of these findings.

The present study used a large sample of offenders to examine
whether there is a consistent pattern of convergence and diver-
gence between INT and EXT criteria and factors of psychopathy
by assessing psychopathy in two different measurement domains
(i.e., self-report, interview and file review) and testing for suppres-
sor situations. Lilienfeld’s Psychopathic Personality Inventory
(PII; Lilienfeld & Andrews, 1996), a self-report measure, and
Hare’s Psychopathy Checklist–Revised (PCL-R; Hare, 1991,
2003), an interview- and file review-based measure, were used to
assess whether relations with latent criteria of INT and EXT
derived from a commonly used self-report inventory of personality
and psychopathy were consistent across these distinct yet
complementary indices of psychopathy.

The Structure of Common Mental Disorders: Latent
Criteria of Internalizing and Externalizing

Factor analytic work using large diagnostic datasets reveals that
common mental disorders in adulthood cohere around higher order
factors of INT and EXT (Krueger, 1999; Krueger, Caspi, Moffitt,
& Silva, 1998; Vollebergh et al., 2001). EXT encompasses the
shared variance among child and adult antisocial behavior along
with alcohol and drug dependence and can be viewed as a broad
vulnerability for disinhibitory psychopathology (Krueger et al.,
2002). Recently, Krueger, Markon, Patrick, Benning, and Kramer
(2007) advanced a hierarchical conceptualization of the external-
izing spectrum that included specification of a general EXT factor
along with separable subfactors related to aggression and sub-
stance use. INT reflects the shared variance between unipolar
depression and anxiety disorders and may be conceptualized as the
general propensity to experience high negative emotionality
(NEM) in terms of fear, sadness, worry, and distress (Mineka,
1998). When estimated as latent factors, INT and EXT tend to be
moderately intercorrelated ($r = .5$; Krueger & Markon, 2006).

In the current work, we used the Personality Assessment Inven-
tory (PAI; Morey, 2007), a self-report measure of personality and
psychopathology, to derive latent criterion variables of INT and
EXT. Prior factor analyses of the PAI scales have yielded factors
analogous to INT and EXT (Frazier, Naugle, & Haggerty, 2006;
Ruiz & Edens, 2008). INT and EXT have traditionally been
conceptualized and measured as latent factors that transcend their
specific indicators and represent shared etiologic processes that
confer a vulnerability to a broad range of psychopathology.

Factors of Psychopathy and Relations with Externalizing
Criteria: The Impact of Method Variance

Indicators of EXT are among the most highly investigated
criteria of psychopathy. The bulk of this research has been con-
ducted with the PCL-R—a 20-item rating scale that is scored based
on review of institutional file data and a semistructured interview
and is designed for use within forensic settings. Although the
appropriate model for conceptualizing the item structure of the
PCL-R remains a matter of debate, the traditional perspective is
that its item content reflects at least two broad factors, which are
commonly measured in psychopathy research (Harpur et al.,
1989).

PCL-R Factor 1 (F1) comprises affective–interpersonal features
from the seminal clinical literature (e.g., superficial charm, egocentricity, manipulativeness, shallow affect). PCL-R Factor 2 (F2) encompasses items reflecting a deviant lifestyle and overt criminality (e.g., poor behavioral controls, impulsivity, irresponsibility, juvenile delinquency).

Although moderately correlated ($r = .50$), PCL-R factor scores
exhibit distinct relations with EXT criteria. Specifically, F2 shows
robust positive associations with symptoms of conduct disorder
and adult antisocial behavior (Hare, 2003; Hart & Hare, 1989),
alcohol and drug abuse or dependence (Reardon, Lang, & Patrick,
2002; Smith & Newman, 1990), and violence (Skeem & Mulvey,
2001), whereas F1 exhibits only modest associations with these
variables. A similar divergence has been observed for personality
traits, with F2 correlating selectively with self-report measures of
disinhibition (e.g., low constraint, low conscientiousness, high
sensation seeking; Hare, 2003; Harpur et al., 1989; Lynam &
Derefinko, 2006; Patrick, 1994; Verona, Patrick, & Joiner, 2001).

Notably, controlling for the shared variance across the PCL-R
factors, associations with EXT criteria can be explained entirely by
F2. In accordance with this research, Patrick et al. (2005) explored
structural relations between the PCL-R factors and a latent EXT
factor defined by symptoms of child and adult antisocial behavior
and substance dependence. In contrast with F1, which exhibited
a nonsignificant negative correlation with EXT after controlling for
F2 scores, the unique variance in F2 was highly correlated with the
latent EXT factor ($r = .94$).

Similar findings have also been reported between EXT and
psychopathy as measured by scores on the PPI. This self-report
measure was designed to capture personality constructs of psy-
chopathy from early clinical conceptions and, in contrast to the

1 Alternative three- and four-factor models have been proposed for the
PCL-R. In the three-factor model (Cooke & Michie, 2001), PCL-R F1 is
parsed into one factor comprising items reflecting an “arrogant and deceit-
ful interpersonal style” and another factor consisting of items reflecting
“deficient affective experience.” The third factor (i.e., impulsive and irre-
sponsible behavioral style) represents the social deviance component of
the PCL-R in terms of the nine items from the original PCL-R F2. The
remaining PCL-R items, which are explicit indicators of antisocial behav-
ior, were omitted from the three-factor model. These items are included as
a distinctive antisocial facet in Hare’s (2003) four-factor model of the
PCL-R, which also includes interpersonal, affective, and lifestyle facets
consisting of items equivalent to the factors of Cooke and Michie’s (2001)
model. Analyses in the current study focused on scores for PCL-R F1 and
F2 because these broad factors have been the focus of most prior research
on relations between the PCL-R and psychopathological criteria, and their
convergent and discriminant validity in this arena is well documente.
PCL-R, does not explicitly reference criminal behavior. Evidence has accumulated that a largely orthogonal two-factor structure, conceptually similar to the two-factor model of the PCL-R, underlies the PPI’s subscales (Benning, Patrick, Hicks, Blonigen, & Krueger, 2003; Patrick, Edens, Poythress, Lilienfeld, & Benning, 2006; Wilson, Frick, & Clements, 1999; Witt, Donnellan, Blonigen, Krueger, & Conger, 2009; but see Neumann, Malterer, & Newman, 2008). PPI-1 (“fearless dominance,” Benning, Patrick, Blonigen, Hicks, & Iacono, 2005) is marked by traits of dominance, narcissism, stress immunity, and fearlessness—core features of the affective–interpersonal factor of psychopathy. PPI-2 (“impulsive antisociality”; Benning, Patrick, Blonigen, et al., 2005) comprises traits of impulsivity, aggression, and alienation—alogous to the social deviance factor of psychopathy. Validity studies indicate that scores on the PPI factors, computed either directly from the PPI or estimated from the Multidimensional Personality Questionnaire (MPQ; Patrick, Curtin, & Tellegen, 2002), yield associations with EXT comparable to the PCL-R factors (Benning, Patrick, Blonigen, et al., 2005; Benning et al., 2003; Edens, Poythress, Lilienfeld, Patrick, & Test, 2008; Patrick et al., 2006; Poythress et al., 2009). For example, example, PPI-2 is positively related to child and adult antisocial behavior, substance abuse or dependence, and disinhibitory personality traits, whereas PPI-1 exhibits negligible to modest associations with these criteria (Benning, Patrick, Blonigen, et al., 2005; Benning et al., 2003; Blonigen, Hicks, Krueger, Patrick, & Iacono, 2005; Patrick et al., 2006).

Despite similarity in these patterns of associations, it is notable that the magnitude of the validity coefficients for the PPI in relation to EXT criteria are modest relative to the PCL-R, thus raising concerns about the extent to which scores on the PPI are associated with indices of deviant behaviors. For example, studies have reported higher correlations between PCL-R scores and symptoms of conduct disorder, antisocial personality disorder, and substance abuse or dependence (e.g., Hare, 2003) than those reported for the PPI (Benning, Patrick, Blonigen, et al., 2005; Blonigen et al., 2005). However, an important issue noted long ago by Campbell and Fiske (1959)—yet often overlooked in the psychopathy literature—is that evaluation of the validity of scores on psychological measures must account for the impact of method variance that is shared (or unshared) across predictors and criteria. From this perspective, PCL-R scores would be expected to be more highly predictive of interview-based assessments of EXT, whereas PPI scores would be likely to converge more strongly with self-report assessments of EXT. The impact of method variance is likely applicable to a range of external criteria beyond EXT and warrants greater consideration when evaluating the criterion and construct validity of psychopathy measures. To this end, our analyses used available criterion indicators of self-report and interview-based measures of EXT to evaluate correspondence with psychopathy scores derived from the same versus different measurement domains.

Factors of Psychopathy and Relations with Internalizing Criteria: The Role of Suppressor Effects

Links between psychopathy and low INT distress are discussed prominently among early clinical conceptualizations of the syndrome. For example, Cleckley (1941) contended that psychopathy was marked by a diminished capacity for emotional distress, as manifested by an absence of nervousness or other psychoneurotic manifestations and low incidence of suicide, suggesting resilience to disorders of anxiety or depression. Moreover, Karpman (1941) identified a subtype of “primary” psychopathy marked by reduced anxiety and emotional distress, and Lykken (1957, 1995) proposed low fear as representing the etiologic core of primary psychopathy. Psychophysiological research provides support for this view, including evidence that individuals with high levels of psychopathic features show attenuated fear-potentiated startle and diminished autonomic reactivity to threatening stimuli (Levenston, Bradley, & Lang, 2000; Patrick, Bradley, & Lang, 1993; Vanman, Mejia, Dawson, Schell, & Raine, 2003; Verona, Patrick, Curtin, Bradley, & Lang, 2004).

In terms of the psychometric evidence, findings to date on the link between psychopathy and low INT have been mixed. In contrast to traditional clinical conceptions, PCL-R total scores exhibit null or in some cases modest positive associations with measures of INT (Hare, 2003; Schmitt & Newman, 1999). At the factor level, some studies have reported negative correlations between F1 scores and indicators of INT (e.g., anxiety, self-criticism; Harpur et al., 1989), whereas others have reported null associations (Schmitt & Newman, 1999; Vitale, Smith, Brinkley, & Newman, 2002). Similar ambiguity is evident for F2, with several studies reporting modest, positive correlations with measures of negative affect (Hale, Goldstein, Abramowitz, Calamari, & Kosson, 2004; Shine & Hobson, 1997; Vitale et al., 2002) and others reporting no relationship (Harpur et al., 1989; Schmitt & Newman, 1999).

In contrast with most published research, which has reported only zero-order correlations, studies controlling for the shared variance between the PCL-R factors using partial correlation or regression methods reveal divergent relations with INT. That is, scores on F1 and F2 exhibit negative and positive associations, respectively, with INT criteria including stress reaction, negative affectivity, and suicide attempts (Douglas et al., 2008; Patrick, 1994; Verona et al., 2005, 2001). Such findings suggest the presence of suppressor effects in the association between PCL-R

2 Like the question of the PCL-R item structure, the issue of the best-fitting structural model of the PPI subscales remains a topic of some debate. Although the PPI two-factor model has been replicated in exploratory factor analyses with different samples (Benning et al., 2003; Benning, Patrick, Salekin, & Leistico, 2005; Ross, Benning, Patrick, Thompson, & Thurston, 2009), Neumann et al. (2008) recently reported that the PPI two-factor model achieved less than adequate fit when evaluated using confirmatory factor analytic (CFA) criteria. However, concerns have been raised in the literature that CFA may be overly conservative as a method for evaluating the fit of structural models of personality inventory data (Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). In addition, the point has been made that the ability of a model to organize relations among constructs is important to consider along with issues of statistical goodness-of-fit (Westen & Rosenthal, 2003). As with the two factors of the PCL-R, an extensive evidence base now exists to support the PPI factors as meaningful and distinct measures (Benning, Patrick, Blonigen, et al., 2005; Benning et al., 2003; Benning, Patrick, Salekin, & Leistico, 2005; Blonigen et al., 2005; Douglas et al., 2008; Edens et al., 2008; Patrick et al., 2006; Poythress et al., 2009; Ross et al., 2009; Sellbom, Ben-Porath, Lilienfeld, Patrick, & Graham, 2003; Uzielle et al., 2007; Witt et al., 2009).
psychopathy scores and INT, specifically cooperative suppression. Cooperative suppression occurs when two correlated predictors evidence opposing relations with a given criterion, such that inclusion of both concurrently in a regression model increases the predictive weight of each with the criterion (Paulhus, Robins, Trzesniewski, & Tracy, 2004). In statistical terms, first-order (beta) coefficients reflecting unique associations for each predictor with the criterion are greater than their zero-order correlations because variance shared with the other predictor is not predictive in the same way, thereby suppressing the association.

In an analysis of relations between the PCL-R factors and facets of NEM (i.e., distress, fearfulness, anger-hostility), Hicks and Patrick (2006) reported evidence of cooperative suppression such that the magnitude of the associations for both F1 and F2 with the NEM facets increased when the two were entered together rather than individually in prediction models. For the PPI, parallel divergences in associations with INT are evident at the zero-order level for the factors of this measure, which are largely independent of one another: PPI-1 is inversely related to self-reports and parental reports of trait anxiety, distress, fear, symptoms of depression and anxiety, and suicide-related behavior; PPI-2 is positively correlated with these criteria (Benning, Patrick, Blonigen, et al., 2005; Benning et al., 2003; Blonigen et al., 2005; Douglas et al., 2008; Patrick et al., 2006; Uzielbo, Verschuer, & Crambeze, 2007; Witt et al., 2009).

If replicable, suppressor effects have important implications for the assessment of psychopathy. Most notably, cooperative suppression suggests that multifaceted constructs embedded within a single instrument that have opposing relations to clinically relevant criteria should be assessed separately from one another to increase their predictive associations with such criteria. In this respect, the PCL-R, although designed as a unitary measure, may exhibit greater predictive power for INT-relevant outcomes (e.g., suicide) when its factors are considered in isolation from one another. By implication, the PPI may be better suited to the prediction of INT-relevant criteria, given that its factors are largely orthogonal rather than moderately correlated as in the PCL-R. Despite these potential implications, suppressor effects have been criticized as being inconsistent, sample-specific, and rarely replicable (Maazen & Baker, 2001; Wiggins, 1973). Hence, further demonstration of their presence in psychopathy measures, particularly in relation to a range of INT criteria, is needed.

**Current Study**

The current study investigated patterns of convergence and divergence with INT and EXT using a multimethod assessment of psychopathy. Relations for psychopathy factor scores were examined in terms of both zero- and first-order associations to test for the presence of suppressor effects on these associations. We predicted that PPI-1 and PCL-R F1 would exhibit weak or negligible associations with EXT and negative associations with INT that would become larger after controlling for (in the case of the PCL-R) shared variance across the psychopathy factors. In contrast, PPI-2 and PCL-R F2 were predicted to correlate positively with both EXT and INT. From an assessment standpoint, the present design represents a notable extension to the existing literature, as it includes multiple assessments of psychopathy in relation to both INT and EXT criteria in the same design, thus permitting evaluation of the impact of method variance on associations between psychopathy and external criteria. To date, no studies have examined relations between psychopathy and INT and EXT across multiple assessment domains.

**Method**

**Participants**

Participants included offenders (prisoners as well as individuals drawn from court-ordered substance use treatment programs) in Florida, Nevada, Oregon, Texas, and Utah. Four of the five treatment programs were community-based; one (Texas) was located within a prison. To mitigate the impact of major mental illness on protocol responses, participants were excluded if they were currently receiving psychotropic medications for active symptoms of psychosis or resided in a mental health unit in prison. Incarcerated participants were deemed eligible if they spoke fluent English and had an estimated IQ ≥ 70 on the Quick Test, a brief screen of intellectual functioning (Ammons & Ammons, 1962). Individuals from substance use treatment programs were required to have completed all detoxification procedures prior to recruitment. At each site, participants were randomly recruited from lists of individuals who met the inclusion criteria. After obtaining informed consent, screening measures for IQ and reading ability were administered, followed by the research protocol for eligible participants.

A total of 1,741 participants were enrolled in the study: 1,413 men (81.2%), 299 women (17.2%), and 29 individuals (1.7%) with missing gender data. The self-reported ethnic and racial composition of the sample was as follows: 1,079 Caucasians (62.0%), 595 African Americans (34.2%), and 67 participants (3.8%) with missing race data. In terms of recruitment site, 911 participants were drawn from prisons (52.3%) and 830 from substance use treatment programs (47.7%).3 The present analyses were based on 1,701 eligible participants with available data on either psychopathy or the INT and EXT criterion measures. The mean age across these participants at time of assessment was 31.64 years (SD = 6.60, range = 17.96–59.37).

**Assessment**

**PCL-R.** Scores on the PCL-R were based on a semistructured interview and review of collateral file information to provide ratings on 20 items. The psychometric properties of the PCL-R have been previously described and include excellent interrater reliability for scores on the PCL-R in research settings (Hare, 2003). PCL-R data were available for 1,572 individuals. Interview protocols were administered by research assistants who received extensive training on the measure, including 2.5 days of face-to-face training from an expert on the PCL-R (Stephen Hart). Prior to the onset of data collection, research assistants completed 10 training tapes and were required to obtain an intraclass correlation (ICC1) of ≥ .80 with criterion ratings. To reduce rater drift, the project coordinator conducted regular site visits over the course of the study to observe PCL-R interviews and independently score interview protocols. For the present sample, interrater reliability

---

3 Supplementary analyses performed separately for criminal offenders and court-referred substance users yielded a pattern of findings in both subsamples that was comparable to the overall sample.
analyses for PCL-R total scores, based on 51 cases derived from these site visits, yielded an intraclass correlation coefficient (ICC; absolute agreement definition; two-way mixed random [raters]/fixed [scale] model) of .88. Scores on PCL-R F1 and F2 were computed by summing scores for items associated with each factor (Hare, 2003). In the current sample, the correlation between scores on PCL-R F1 and F2 was .50.

**PPI.** Self-report scores on psychopathy were assessed with the PPI, which consists of 187 items, endorsed on a 4-point Likert scale (1 = false, 2 = mostly false, 3 = mostly true, 4 = true). The inventory provides a total score as well as scores on eight subscales, which yielded the following internal consistency estimates (Cronbach’s alphas): Machiavellian Egocentricity = .89, Social Potency = .86, Fearlessness = .85, Coldheartedness = .78, Impulsive Nonconformity = .73, Blame Externalization (formally called Alienation) = .87, Carefree Nonplanfulness = .84, Stress Immunity = .80. Satisfactory internal consistency has also been reported for PPI total and scale scores in undergraduate (Lilienfeld & Andrews, 1996), community (Blonigen, Carlson, Krueger, & Patrick, 2003), and offender samples (Neumann et al., 2008; Sandoval, Hancock, Poythress, Edens, & Lilienfeld, 2000; Uezieblo et al., 2007). PPI data were available for 1,607 participants.

Seven of the eight PPI subscales typically cohere into two largely orthogonal factors (PPI-1 & PPI-2; Benning et al., 2003; Patrick, et al., 2006; Witt et al., 2009). Scores for PPI-1 were computed by taking the average of the summed z-scores of Social Potency, Fearlessness, and Stress Immunity (Benning et al., 2003). Scores for PPI-2 were computed by taking the average of the summed z-scores of Machiavellian Egocentricity, Impulsive Nonconformity, Blame Externalization, and Carefree Nonplanfulness. The PPI Coldheartedness scale does not load highly on either PPI factor and was therefore excluded from these computations. Scores on the two factors of the PPI were correlated to a negligible, albeit statistically significant, degree in our large sample (r = -.06, p = .013).

In the current sample, the zero-order correlation between total scores on the PPI and the PCL-R was .43 (p < .01). With regard to corresponding factors of the two instruments, PPI-1 was correlated more strongly with PCL-R F1 than with F2 (rs = .26 and .16, respectively; ps < .01), whereas PPI-2 was correlated more highly with PCL-R F2 than F1 (rs = .38 and .17, respectively; ps < .01). When examining partial correlations to control for the shared variance between the PCL-R factors, PPI-1 was selectively associated with PCL-R F1 (partial r = .21, p < .01) and unrelated to PCL-R F2 (partial r = .04, ns), whereas PPI-2 was preferentially related to PCL-R F2 (partial r = .35, p < .01) and unrelated with PCL-R F1 (partial r = -.03, ns).

**Self-report measures of internalizing & externalizing.** The PAI, a 344-item self-report inventory of personality and psychopathology, was used to measure factor scores of INT and EXT. The PAI consists of 11 clinical scales, five treatment consideration scales, two scales of interpersonal style, and four validity scales. The scales of primary interest in this study included the clinical scales of Anxiety (ANX; α = .91), Anxiety-Related Disorders (ARD; α = .81), Depression (DEP; α = .89), Antisocial Features (ANT; α = .85), Alcohol Problems (ALC; α = .94), Drug Problems (DRG; α = .89), and the treatment consideration scale of Aggression (AGG; α = .91). Numerous studies, including research conducted with criminal offenders, support the construct validity of each of these scales (e.g., Edens & Ruiz, 2008). These scales were chosen a priori based on the following criteria: (a) they represent specific indicators of either INT or EXT according to the literature on the structure of psychopathy (Krueger, 1999; Vollebergh et al., 2001), and (b) they loaded strongly on these dimensions in a recent analysis of PAI scales in offender samples (Ruiz & Edens, 2008). The PAI is appealing for use in forensic and correctional settings due to its relatively low reading level (fourth grade), brevity, solid psychometric properties (Edens & Ruiz, 2005; Morey, 2007), and multiple validity scales. Results herein include PAI data for 1,642 participants, including 39 participants whose scores on the validity scales exceeded cutoffs recommended by Edens and Ruiz (2005) as indicating potentially invalid profiles (i.e., Inconsistency > 79 or Infrequency > 79). These participants were included in the final analyses, as the results did not differ as a function of whether these subjects were included versus excluded. The distributions of several PAI scales were somewhat skewed in the present data. Thus, scale scores were transformed using Blom’s (1958) rank transformation method to normalize the distributions.

**Interview measures of externalizing:** Diagnostic symptom counts of conduct disorder (CD; n = 1,515) and adult antisocial behavior (AAB; n = 1,543) were obtained using the Antisocial Personality Disorder (ASPD) module of the Structured Clinical Interview for DSM-IV Axis-II Personality Disorders (SCID-II; First, Gibbon, Spitzer, Williams, & Benjamin, 1996). This module, based on criteria from the Diagnostic and Statistical Manual of Mental Disorders (4th ed.), yields dimensional and categorical (diagnostic) scores for both CD and ASPD. In this study, the interrater reliability was high for both diagnoses of ASPD (κ = .74; n = 50) and total symptom count (ICC = .86; n = 46), along with similarly high internal consistency (α = .83).

**Data Analysis**

Confirmatory factor analysis (CFA) was used to test the fit of a two-factor model of INT and EXT, ANX, ARD, and DEP were selected as indicators of INT, and ANT, AGG, DRG, and ALC were chosen to estimate EXT. This analysis was conducted with the raw (Blom-transformed) data using maximum likelihood estimation, which accounts for missing data, in the statistical package Mplus 4.1 (Muthén & Muthén, 2006). Evaluation of model fit was based on the chi-square fit statistic (χ²), the comparative fit index (CFI), the Tucker-Lewis Index (TLI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). Chi-square and RMSEA yield overall estimates of model fit, with RMSEA providing an index of the discrepancy in model fit per degree of freedom. SRMR provides an estimate of the average discrepancy between the observed and model-estimated sample statistics. For CFI and TLI, values approaching 1.0 indicate a good fitting model; values ranging from .90 to .95 suggest acceptable fit (Brown, 2006). For RMSEA and SRMR, values less than .08 indicate an acceptable fit to the data (Brown, 2006), whereas values less than .06 suggest a close fit to the data (Hu & Bentler, 1999).

To investigate relations between self-report (PPI) and interview-and file review-based (PCL-R) assessments of psychopathy and INT and EXT, we began with a structural model in which INT and EXT were modeled as latent (correlated) criterion variables with psychopathy scores included as predictors in the model. PPI and PCL-R total and factor scores were modeled as manifest variables
and were entered individually as predictors to assess zero-order associations with the latent criteria. For the PCL-R, scores for the two factors were entered jointly into the model to assess the unique relations between each factor and INT and EXT after controlling for shared variance with the other factor. To evaluate the significance of suppressor effects, the Sobel test statistic (z) was calculated using a publicly available macro (Preacher & Leonardelli, 2006). This test, commonly used in the intervening variable framework of mediation models, provides an index of the significance of the change in the beta coefficient for a predictor variable after entry of a second predictor into the model. Changes in the squared multiple correlation (ΔR²) were used to evaluate the incremental validity of scores on each PCL-R factor over the other in the prediction of the criterion variables.

Our approach of measuring psychopathy at the manifest level is consistent with Hicks and Patrick (2006) in their examination of associations between the PCL-R and latent indices of NEM. Moreover, adjudicating the best-fitting structural model of the PPI and PCL-R—isues of ongoing debate in the literature—is beyond the scope of the current study. By contrast, there is general consensus among scholars that INT and EXT represent latent dimensions of psychopathy that transcend their specific indicators. Thus, we chose to derive an INT and EXT at the latent level to obtain indices of these criteria that are equivalent to their conceptualization and measurement in extant research.

Following the SEM analyses, we evaluated the impact of measurement error on the magnitude of associations between measures of psychopathy and external criteria. Using self-report and interview composites of EXT, we compared zero- and first-order relations with the self-report (PPI) and interview- and file-review-based measures of psychopathy (PCL-R). In contrast to the range of EXT indicators available in self-report PPI, these analyses were limited to antisocial indicators of EXT, as the interview-based (SCID) counterparts were not available for either substance abuse indicators of EXT or for any indicators of INT.

Results

Derivation of Latent Internalizing and Externalizing Criteria

A correlated two-factor model was fit to the PAI scales using CFA, with INT measured by ANX, ARD, and DEP, and EXT measured by ANT, AGG, DRG, and ALC. This model yielded the following fit indices: χ²(13, N = 1,642) = 375.19, p < .01; CFI = .92; TLI = .87; RMSEA = .130; SRMR = .060. Although CFI and SRMR suggested adequate fit, RMSEA indicated marginal fit. A revised model was tested that expanded the two-factor model to include correlated residual terms for pairs of conceptually related indicators of EXT (i.e., ANT and AGG; DRG and ALC). The coherence between these indicator pairs is consistent with recent factor analytic work demonstrating distinct aggression and substance use subfactors in conjunction with an overarching EXT factor (Krueger et al., 2007). This revised model provided a good overall fit to the data, χ²(11, N = 1,642) = 123.64, p < .01; CFI = .98; TLI = .95; RMSEA = .079; SRMR = .040, and a significant improvement in fit compared with the two-factor model without correlated residuals: Δχ²(2, N = 1,642) = 251.55, p < .01. By contrast, inclusion of correlated residuals for indicators of INT attenuated rather than improved the fit of the model. Figure 1 shows the parameter estimates for the correlated two-factor model.

Structural Relations Between Psychopathy Factors and Internalizing and Externalizing

Parameter estimates of relations between the psychopathy factors and INT and EXT criteria are provided in Table 1. Ranges for the fit statistics from these models were as follows: CFI = .94—.95, TLI = .90—.91, RMSEA = .09—.11, SRMR = .049—.066. For INT, a small to negligible association emerged between this factor and overall scores on the psychopathy inventories: PCL-R total scores were uncorrelated with INT; PPI total scores exhibited a small positive association with this criterion. For EXT, a positive association was observed for overall scores on both psychopathy measures: PPI total scores were highly correlated with EXT; PCL-R total scores were moderately correlated with this criterion.

At the level of the psychopathy factors, a divergence in associations was evident for INT versus EXT. For INT, the psychopathy factors evinced opposing associations such that PPI-I was highly negatively related to INT, whereas PPI-2 was highly positively related to INT. A similar pattern emerged for the PCL-R: F1 was negatively correlated with INT; F2 was positively correlated with this criterion. In contrast to the PPI factors, zero-order correlations between the PCL-R factors and INT, although significant, were small in magnitude. However, when overlapping variance between the PCL-R factors was accounted for in the model, the opposing associations with INT increased in magnitude, thus demonstrating cooperative suppression (such analyses were not conducted with the PPI because its factors share almost no variance with one another). Based on the Sobel test, significant increases were observed in the associations with INT for both PCL-R F1 (z = 6.06, p < .01) and F2 (z = −6.44, p < .01) after inclusion of the other factor in the model. Moreover, based on changes in R², both F1

---

4 To evaluate the stability of the INT–EXT model as parameterized in the main analyses, we randomly divided the overall sample in half and tested the fit of the INT–EXT model (including correlated residuals) in each subsample. The model fit well in both half-samples, with the magnitude of correlated residual terms similar in each model: For ANT and AGG, the magnitude of the correlated residual was .37 (p < .001) in both subsamples; for ALC and DRG, it was .06 in one half-sample and .08 in the other (ps = .056 and .013, respectively).

5 As emphasized by Marsh, Hau, and Wen (2004), the stringent cutoff values for goodness-of-fit (GOF) indices proposed by Hu and Bentler (1999) are more effective at identifying model misspecification based on differences in nested models, and they cautioned that these values should be viewed as “rules of thumb” rather than “golden rules.” Furthermore, as argued by Westen and Rosenthal (2003) as well as Marsh et al. (2004), statistical GOF is merely one criterion from which to evaluate the appropriateness of a given structural model. From this perspective, although some of the fit statistics from the structural models of INT and EXT that included the psychopathy factors as predictors were marginal by the Hu and Bentler (1999) standards, the model estimates were comparable when modeling INT and EXT as observed scales, and were highly consistent with past research.
\( \Delta R^2 = .03, p < .01 \) and F2 \( \Delta R^2 = .03, p < .01 \) demonstrated incremental validity over the other in the prediction of INT.

For EXT, a different pattern emerged relative to INT. PPI-2 was highly correlated with EXT, whereas PPI-1 was only weakly, albeit significantly, related to this criterion. For the PCL-R, F2 also showed a moderate positive association with EXT in terms of both zero-order and first-order coefficients, whereas zero-order and first-order coefficients for PCL-R F1 were negligible and modestly negative, respectively. Sobel tests indicated a significant decrease in the association with EXT for PCL-R F1 \( z = 13.13, p < .01 \) and a significant increase in the association with EXT for PCL-R F2 \( z = -3.39, p < .01 \) after inclusion of the other factor in the model. In addition, the change in \( R^2 \) was significant for both PCL-R factors when included as predictors of EXT; however, the magnitude of incremental prediction was substantially larger for F2 \( \Delta R^2 = .26, p < .01 \) than F1 \( \Delta R^2 = .01, p < .01 \).

### Table 1

**Structural Relations Between PPI & PCL-R Total and Factor Scores and Latent Internalizing and Externalizing Criterion Variables**

<table>
<thead>
<tr>
<th>Latent criterion variable</th>
<th>PPI Zero-order (r)</th>
<th>PCL-R</th>
<th>PPI-1</th>
<th>PPI-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internalizing</td>
<td>.14**</td>
<td></td>
<td>-51**</td>
<td>.52**</td>
</tr>
<tr>
<td>Externalizing</td>
<td>.74**</td>
<td></td>
<td>.6**</td>
<td>.79**</td>
</tr>
</tbody>
</table>

Note. PPI = Psychopathic Personality Inventory; PCL-R = Psychopathy Checklist-Revised. Zero-order (r) = correlations between psychopathy factors and Internalizing and Externalizing when evaluated as single predictors in the model. First-order (\( \beta \)) = beta weights for psychopathy factors when entered together as predictors of Internalizing and Externalizing in the model.

**p < .01.

### Relations Between Psychopathy Factors and Externalizing by Measurement Domain

A comparison of effects for the PPI versus the PCL-R in Table 1 reveals that correlations for the PPI factors with INT and EXT were markedly higher than the corresponding correlations for the

6 We supplemented the analyses of relations between PCL-R psychopathy and INT and EXT in Table 1 by parsing each of the PCL-R factors into narrower facet scores based on Hare’s (2003) four-factor model, which encompasses the three factors in Cooke and Michie’s (2001) model. Zero-order coefficients were computed, along with first-order coefficients in which the unique predictive association for each PCL-R facet (F1: interpersonal, affective; F2: lifestyle, antisocial) was examined after controlling for variance associated with the complementary factor (F1 or F2). These analyses revealed that (a) the negative association of PCL-R F1 with INT (revealed most clearly after variance associated with F2 was controlled for) was attributable to both its interpersonal and affective facets; (b) the positive association of PCL-R F2 with INT (most evident after controlling for variance associated with F1) was attributable to its lifestyle facet; (c) the positive association of F2 with EXT (robust across the zero- and first-order coefficients) was attributable to both its lifestyle and antisocial facets; and (d) the small negative association of F1 with EXT (evident only in the beta coefficient after variance associated with F2 was controlled for) was attributable to its affective facet (cf. Poullides, Dembo, Wareham, & Greenbaum, 2006). Further details of these analyses are available from Daniel M. Blonigen upon request.
PCL-R factors. This difference is potentially explainable by the fact that scores for the PPI were derived from the same measurement domain as the PAI, whereas scores for the PCL-R were derived from a different measurement domain. Table 2 addresses this issue by comparing correlations for the PPI and PCL-R with EXT composite scores derived from self-report indicators (i.e., PAI ANT and AGG scales) versus interview indicators (i.e., child and adult symptoms of ASPD assessed using the SCID). EXT composite scores were computed as the average of standardized (z) scores for indicators from one or the other assessment domain. For both the PCL-R and PPI, correlations with EXT were uniformly higher when assessed within the same measurement domain: PPI scores, particularly total scores and PPI-2, exhibited significantly higher correlations with the self-report EXT composite than with the interview-based EXT composite; PCL-R total and factor scores demonstrated significantly larger correlations with the interview-based EXT composite than with the self-report EXT composite. Relative to the zero-order coefficients, the first-order coefficients showed a similar pattern of method convergence. With the exception of PPI-1, all parameter estimates were significantly different across measurement domains within each predictor. Notably, the magnitude of association for scores on the two psychopathy instruments with EXT was highly comparable when EXT scores were derived from the same measurement domain, and when shared variance between the PCL-R factors was removed. When the self-report and interview composite scores were combined into a single index of EXT (i.e., mean of the composites), scores on this combined composite were also correlated with PCL-R F2 (r = .65, β = .63) and PPI-2 (r = .65) to a comparably high level.7

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Relations Between Psychopathy Total and Factor Scores and Externalizing Composites by Measurement Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion variable</td>
<td>Zero-order (r)</td>
</tr>
<tr>
<td></td>
<td>PPI total</td>
</tr>
<tr>
<td>Self-report ANT/AGG composite</td>
<td>.70**</td>
</tr>
<tr>
<td>Interview CD/AAB composite</td>
<td>.49**</td>
</tr>
</tbody>
</table>

Note. PPI = Psychopathic Personality Inventory; PCL-R = Psychopathy Checklist-Revised; ANT = PAI-Antisocial Features; AGG = PAI-Agression; CD = conduct disorder symptoms based on the Structured Clinical Interview for DSM-IV Axis-II Personality Disorders (SCID); AAB = SCID-based adult antisocial behavior symptoms. Composites are observed variables and represent average z-scores of respective scales. Bold = correlations within the same assessment domain. Italic = correlations significantly different across measurement domains within each predictor. ** p < .01.

For the self-report EXT composite, Sobel tests indicated a significant decrease in the association with this criterion for PCL-R F1 (z = 14.07, p < .01; ΔR² = .17, p < .01) and a nonsignificant change in the relationship with this criterion for PCL-R F2 (z = -1.50, ns; ΔR² = .001, ns) after inclusion of the other factor in the model. For the interview-based EXT composite, Sobel tests indicated a significant decrease in the relationship with this criterion for both PCL-R F1 (z = 17.62, p < .01) and PCL-R F2 (z = 5.34, p < .01). Consistent with findings presented in Table 1, the magnitude of incremental prediction for this criterion was substantially larger for F2 (ΔR² = .32, p < .01) than F1 (ΔR² = .01, p < .01).

Discussion

The current study investigated the impact of method variance and suppressor effects in the relationship between factors of psychopathy and INT and EXT criteria operationalized via latent variables. Consistent with prior research, scores on the affective interpersonal factor of psychopathy (PCL-R F1, PPI-1) were negatively associated with INT, whereas scores on the social deviance factor (PCL-R F2, PPI-2) were positively associated with INT. In the case of the PCL-R, clear evidence of cooperative suppression was obtained such that associations with INT increased significantly when both factors were included concurrently as predictors of this criterion compared to when either was considered alone as a predictor. This replicates prior work demonstrating suppressor effects in correlations for the PCL-R factors (or factors of cognate psychopathy measures) with measures of negative affect (Frick et al., 2000, 1999; Hicks & Patrick, 2006; McHoskey, Worzel, & Szyarto, 1998) and suicidal tendencies (Douglas et al., 2008; Verona et al., 2005). With respect to EXT, scores for both PCL-R F2 and PPI-2 were positively correlated with this criterion, adding to a growing body of data establishing a close correspondence between the social deviance factor of psychopathy and disinhibitory psychopathology (Blonigen et al., 2005; Patrick et al., 2005; Smith & Newman, 1990).

Moderating Impact of Method Variance: Implications for Evaluating the Validity of Psychopathy Measures

Method variance is a crucial issue that tends to be overlooked in studies evaluating the criterion and construct validity of psychopathy measures. In their classic article, Campbell and Fiske (1959) pointed out that compared with measures of the same construct assessed within the same methodological domain, which are expected to show high correlations with one another (.7 to .9), alternative measures of the same construct assessed in different methods are expected to show lower correlations (e.g., .5 to .6). A potential alternative interpretation is that divergence in the correlations for the same versus different measurement domains could have reflected differences in the constructs assessed rather than the domain in which they were measured. That is, self-report indicators of EXT (PAI-based ANT and AGG) may have indexed a different construct than the interview-based indicators (SCID-based CD and AAB). One piece of evidence arguing against this possibility is that the correlation between the self-report and interview-based EXT composites (r = .53) was comparable to the correlations between indicators within each measurement domain (i.e., CD and AAB were correlated, .48; ANT and AGG were correlated, .61).
methodological domains are expected to show only moderate levels of association (.4 to .6). Indeed, the current finding that the level of association with the self-report criterion variables for the PCL-R factors was lower than that for the corresponding PPI factors highlights the importance of method variance as a moderator of relations between psychopathy factors and external criteria.

The present work addressed this issue by comparing the relative strength of associations for scores on the PCL-R versus the PPI with composite scores of EXT based on interview and self-report indicators. As predicted, the magnitude of the associations was comparably high for the two instruments when the predictor and criterion were assessed in the same measurement domain and comparably moderate when the predictor and criterion were assessed in different domains. Remarkably, when derived from the same measurement domain and when the shared variance between the PCL-R factors was controlled for, the magnitude of the associations with EXT for the two psychopathy measures were nearly identical (i.e., zero-order coefficients for PPI-1 and PPI-2 with self-report based EXT were .15 and .69, respectively; first-order coefficients for PCL-R F1 and F2 with interview-based EXT were .11 and .65, respectively). These results can be tied to findings from existing published work revealing stronger associations with clinical ratings of ASPD symptoms for the PCL-R (Hare, 2003) than the PPI (Benning, Patrick, Blonigen, et al., 2005; Blonigen et al., 2005). Collectively, the data suggest that comparisons of the validity of interpretations based on self-report and interview-based measures of psychopathy can be facilitated by combining measures from different assessment domains into composite criterion variables (see Hare & Cox, 1978).

An additional benefit of our multimethod approach to psychopathy assessment is that it helped to clarify the degree of correspondence between the constructs underlying factors of psychopathy measures and criteria of INT and EXT. For example, in contrast to past findings revealing moderate correlations between scores on PPI-2 and interview-based symptoms of EXT (Blonigen et al., 2005), the present observation of a strong overlap between scores on PPI-2 and PAI-based EXT is consistent with the notion that these two variables represent alternative measures of the same construct within the same measurement domain. In conjunction with other work demonstrating high correspondence between PCL-R F2 and EXT when scores on the latter incorporate interview measures (Patrick et al., 2005), the current findings suggest that PCL-R F2 and PPI-2 are measuring nearly equivalent constructs within different domains of assessment. By implication, the modest intercorrelations previously reported between these corresponding factors of the PPI and PCL-R may be due, in part, to the influence of method variance.

Although the current study did not include interview-based indicators of INT, by extension it can be hypothesized that differences in the strength of the negative associations with INT for PCL-R F1 ($\rho = -.12/-21$) versus PPI-1 ($r = .51$) is at least partially attributable to method variance. Consistent with this interpretation, Blonigen et al. (2005) reported significant correlations between MPQ-estimated PPI-1 scores and interview-based symptoms of INT (i.e., -.26 and -.25 for women and men, respectively). These correlations were lower than the correlation between INT and PPI-1 in the current study, but comparable to the association between INT and the unique variance in PCL-R F1. Thus, the negative association between PCL-R F1 and scores on INT would likely be amplified for a composite that incorporated interview-based indicators of this criterion.

However, the fact that the negative relationship between PAI-based INT and PPI-1 was moderate rather than high in magnitude indicates that these two variables are predicting related rather than equivalent constructs. Furthermore, across the PPI and PCL-R, the strength of the positive association between scores on the social deviance factors of these measures and EXT exceeded the strength of the negative association between scores on the affective–interpersonal factors and INT. These differences in relative strength suggest that scores on the social deviance factors are more directly indicative of EXT tendencies than scores on the affective–interpersonal factors are of the absence of INT tendencies. In other words, the social deviance factor of psychopathy is largely synonymous with the EXT dimension of psychopathy, whereas the construct measured by scores on the affective–interpersonal factor only partly reflects resilience against INT symptomatology.

**Moderating Impact of Suppressor Effects: Implications for Theory and Clinical Assessment**

Beyond the role of method variance, the present study elucidates another source of difference between the PPI and PCL-R in their relations with external criteria. In contrast to the factors of the PCL-R, scores on PPI-1 and 2 are largely independent. As a function of this, differential associations between the PPI factors and criterion variables are clearly revealed in simple (zero-order) correlations, whereas distinctive relations with many criterion measures for the PCL-R factors emerge more clearly when their shared variance is accounted for using regression or structural analytic techniques. The current data helps to reconcile observed differences in criterion-related validity across the PPI and PCL-R as well as disparities between theory and empirical observations for the PCL-R. Specifically, the identification of cooperative suppressor effects for the PCL-R factors in relation to INT reconciles the strong emphasis that Cleckley (1941) placed on “absence of nervousness or psychoneurotic manifestations” (p. 339) with the null or slight positive association typically reported between total scores on the PCL-R and indices of anxiety and depression (Hare, 2003).

From the standpoint of clinical assessment, the presence of suppression (particularly cooperative suppression) suggests that distinguishable constructs are embedded within a putatively unitary measure. Consequently, instruments underpinned by multifaceted constructs with opposing relations to clinically relevant criteria should be assessed separately from one another to identify individuals who are high on one factor but low on the other, thereby increasing their ability to predict such criteria. In this respect, the PCL-R, although designed as a unitary measure, may have greater utility in predicting INT-relevant outcomes when its factors are measured in isolation. For example, Hicks and Patrick (2006) noted that separate measurement of the PCL-R factors could significantly improve prediction of clinical outcomes marked by negative affect (e.g., aggression, suicide, treatment responsiveness). By comparison, the PPI may be better suited to predict these criteria given that its factors are largely orthogonal and thus index distinctive components of the psychopathy construct in a more clearly differentiated way than the factors of the PCL-R. This probably reflects the fact that the PPI was developed...
to assess the spectrum of psychopathy-related traits without requiring that these traits cohere together as a unitary dimensional entity. In sum, from an assessment standpoint, the predictive utility of the PPI and PCL-R can be maximized via separate assessment of the factors of these measures, which in turn would serve to retain information that is obscured using total scores (Patrick et al., 2006).

Limitations and Future Directions

A key limitation of the current study is that interview-based indicators of EXT were limited to CD and AAB, and interview-based indicators of INT were unavailable. This issue is readily addressable in future work by administering modules from the SCID that assess for anxiety, mood, and substance use disorders along with self-report and interview measures of psychopathy. In conjunction with parallel self-report measures of INT and EXT, this methodological approach would provide for a complete same-versus cross-domain analysis of relations between factors of psychopathy and broad factors of psychopathology and would allow for a direct assessment of the proportion of variance among the measures that is attributable to method variance (i.e., an SEM-based multitrait-multimethod analysis).

An additional limitation was that analyses focused exclusively on data from adult offenders. Future studies are needed to extend these results to different developmental time points. For example, it has long been known that common disorders of childhood are also organized around broad INT and EXT dimensions (Achenbach & Edelbrock, 1984). Thus, relations between these psychopathological criteria and factors of psychopathy may be investigated in youth samples using inventories designed to measure psychopathic features in these populations (e.g., Andershed, Kerr, Stattin, & Leander, 2002; Frick & Hare, 2001; Lynam, 1997). Including multiple measures of psychopathy and psychopathology in alternative assessment domains (e.g., parent ratings versus self-report) at different stages of the lifespan would provide a valuable extension of the impact of method variance and suppressor effects on the magnitude of observed associations.

Finally, future efforts are needed to clarify points of convergence and divergence between alternative self-report and PCL-R-based measures of psychopathy (Poythress et al., 2009). For example, a key conclusion emerging from the current work is that although PCL-R F2 and PPI-2 measure nearly equivalent constructs, PCL-R F1 and PPI-1 index related but nonidentical constructs. Research aimed at elucidating commonalities and distinctions between the affective-interpersonal factors of these two inventories (with consideration of issues of method variance and suppressor effects) stands to advance our understanding not only of what comprises each measure but also, more generally, what range of distinctive constructs comprise the syndrome of psychopathy (Patrick, Fowles, & Krueger, 2009).

References


izing and externalizing dimensions within the Personality Assessment Inventory. *Journal of Personality Assessment*, 90, 585–592.


Received January 3, 2009
Revision received June 25, 2009
Accepted July 1, 2009