The Perils of Partialling

Cautionary Tales From Aggression and Psychopathy

Donald R. Lynam
Purdue University

Rick H. Hoyle
Duke University

Joseph P. Newman
University of Wisconsin–Madison

*Although a powerful technique, the partialling of independent variables from one another in the context of multiple regression analysis poses certain perils. The present article argues that the most important and underappreciated peril is the difficulty in knowing what construct an independent variable represents once the variance shared with other independent variables is removed. The present article presents illustrative analyses in a large sample of inmates (*n = *696) using three measures from the psychopathy and aggression fields. Results indicate that in terms of relations among items on a single scale and relations between scales, the raw and residualized scores bore little resemblance to one another. It is argued that researchers must decide to which construct—the one represented by the original scale or the one represented by the residualized scale—conclusions are meant to apply. Difficulties in applying the conclusions to the residualized scale are highlighted and best practices suggested.*

Keywords: psychopathy; aggrression; partialling; suppression

The partialling of independent variables from one another in the context of multiple regression is a powerful and widely used procedure. This procedure involves removing the variance from one variable that is shared with others in a set, allows the examination of the relations between a set of independent variables and a dependent variable, and makes the testing of predictive and theoretical models possible. Results from such an analysis are very informative. They tell how well a group of independent variables predicts a dependent variable, they reveal which independent variable makes the largest incremental contribution to the prediction of the dependent variable after taking into consideration the remaining independent variables, and they indicate whether the independent variable of interest remains related to the dependent variable after holding other independent variables constant. In short, multiple regression allows researchers to test some of their most interesting hypotheses.

But as with any statistical procedure, perils and pitfalls need to be avoided if confident conclusions are to be drawn. One of the most widely known problems is that of multicollinearity, the substantial correlation among a set of independent variables. In writing about this problem, Cohen and Cohen (1983) indicated that multicollinearity gives rise to three distinct problems: the substantive interpretation of partial coefficients, sampling stability, and computational accuracy. Although the problem with

This study was supported by grants from the National Institute of Mental Health to Joseph P. Newman (MH53041) and Donald R. Lynam (MH60104). We wish to acknowledge the indispensable cooperation of the staff at the Oakhill and Columbia Correctional Institutions. Correspondence concerning this article should be addressed to Donald R. Lynam, Department of Psychological Sciences, Purdue University, 703 Third St., West Lafayette, IN 47907-2081; e-mail: dlynam@psych.purdue.edu.

Assessment, Volume 13, No. 3, September 2006 328-341
DOI: 10.1177/1073191106290562

computational accuracy has been virtually solved, the other two problems remain.

The issue regarding sampling stability has received the lion's share of attention. This problem is often couched in terms of tolerance, the proportion of variance in one independent variable not accounted for by other independent variables, or the variance inflation factor, 1.0 divided by the tolerance. As multicollinearity increases, tolerance decreases, the variance inflation factor increases, and confidence intervals for the coefficients increase, leading to highly unstable partial coefficients.[1] This problem has been well-addressed and there are explicit recommendations regarding how much collinearity is too much collinearity in terms of tolerance and the variance inflation factor (e.g., Morrow-Howell, 1994).

The third potential pitfall to partialling, the substantive interpretation of partialled variables, has been raised by researchers in various fields. Writing for sociologists, Gordon (1968) noted several specific ways in which partial regression coefficients might be misleading. He noted that "although warnings concerning multicollinearity are to be found in statistics text, they are insufficiently informative . . . because the problem is essentially one of substantive interpretation rather than mathematical statistics" (p. 592). In another classic piece, Meehl (1975) raised important questions about whether so-called nuisance variables should be partialled out in multivariate analyses. Similarly, authors have railed against the use of analysis of covariance to equate groups on preexisting differences (see Chapman & Chapman, 1973; Lord, 1967). The crux of the problem identified by almost all critics is that it is difficult to know what construct an independent variable represents once the variance shared with other independent variables is removed. Miller and Chapman (2001), in their recent article on misuses of analysis of covariance, argued that the substantive problem with using analysis of covariance to correct for preexisting group differences is that one does not know what is left in the grouping variable (Grp) once the variance shared with the covariate (Cov) is removed (Grp_{res}). They offered that

> the central problem is that often one does not know what Grp_{res} represents when Cov and Grp are related. The grouping variable, its essence, has been altered in some substantive way that is frequently not specifiable in a conceptually meaningful way. Thus, Grp_{res} is not a good measure of the construct that Grp is intended to measure. (p. 43)

We believe that the concern regarding what is left in a variable applies any time one variable is partialled from another variable.

If researchers simply want to make statements about the relative contribution of a scale in predicting a specific outcome, difficulties with substantive interpretation of partial coefficients are unlikely. This, however, is not the kind of conclusion that many wish to draw. More often, researchers want to make statements about the constructs being measured by the original scales, that is, they want to flesh out the nomological networks surrounding their constructs of interest. This is a more complex question in which the substantive interpretation of the partial coefficients is critical. Partialling changes variables. The degree to which a partialled variable differs from the original variable depends on a number of factors, including the degree of overlap among variables, the reliability of the variable, and the degree of internal consistency of the variable, all of which contribute to the theoretical tightness of the variables involved and the interpretation of the partial coefficients.

Degree of Overlap

As the degree of overlap between the variables to be partialled from one another increases, the similarity between the raw and residualized scores decreases. This can be seen by examining the correlation between raw and residualized scores or the proportion of variance removed from a score as overlap between the variables increases. For example, the correlation between the raw and residualized score is almost unity ($r = .995$) when the correlation between scales is .10, but it decreases substantially as the correlation between raw scores increases to .866 for a correlation of .50 and .316 for a correlation of .90.[2] In terms of the proportion of variance removed, when variables are correlated at .10, only 1% of the variance is removed. As the correlation between raw scores increases, so does the proportion of variance removed; this proportion is 25% for a correlation of .50 and 81% for a raw-raw correlation of .90. Of importance, the variance removed typically comes entirely out of the reliable variance of the measure; hence, these proportions must always be understood in terms of the reliabilities of the measures.

Reliability

In addition to influencing the interpretation of what is removed in the partialling process, reliability has implications for what remains. The variation in a set of scores can be decomposed into two variances: the variance attributable to the target construct and the variance attributable to errors. It is important to note that random error and systematic error not shared by the variables involved in partialling remains in the residualized score and now comprises a larger part of the variable, that is, the residualized score is less reliable than the raw score. For

example, if the reliability of the raw scores is .80 and the correlation between the raw scores is .50, then the effective reliability of the residualized scores is .73.[3] If the reliability of the raw scores is .80 and the correlation between the two variables is .75, a value more typical of the analytic situation we describe later, then the effective reliability of the residualized scores drops to .54.

Internal Consistency

Although difficult to quantify, the contribution of heterogeneity among items in a measure or, conversely, internal consistency of items is important as well. Heterogeneous measures run the risk of greater dissimilarity following partialling than more homogeneous measures. This is because the partialling process may remove variance that is specifically associated with some elements of the measures but not others.

Theoretical Tightness

All of these psychometric characteristics contribute to the theoretical tightness and interpretability of the variable. Narrow, reliably measured, homogeneous variables are theoretically tight in the sense that the content of the variable is known and its boundaries are well-defined. When theoretically tight variables are partialled from one another, it is easier to know what is left in the residualized variables. To the degree that the variables are unreliable and heterogeneous, they are looser and more ill-defined. In these cases, it is difficult to know what is left when one variable is partialled from the other, and the inconsistency of zero-order and partial relations can be expected to be greater.

Changes in the Nomological Network After Partialling

Variables are known in terms of their internal structures and the nomological networks in which they are embedded, that is, the constructs measured by scales are understood by reference to the content of the items comprising the scale and by the correlations of scores on the scale with scores on other scales. Unfortunately, the internal structure and the external correlations can vary considerably, particularly in the case of highly correlated, less reliable, heterogeneous variables, following partialling. In fact, the relations may vary to such a degree that results from analyses that involve partialling cannot be applied to the original constructs that inspired the measures of the variables. Of course, the degree of the difficulty in applying the results from the partial analyses to the original construct depends entirely on the nature and degree of change in the internal structure of the variable and its nomological network. Some changes are unlikely to be problematic (e.g., across the board decreases), whereas others will prove more nettlesome (e.g., suppression).

The simplest and least problematic change in the nomological network is when the relations between the variable of interest and the correlates that comprise its nomological network (validation measures) decrease across the board following partialling. Imagine that a variable that is moderately correlated with each of a set of variables at the zero-order level becomes less strongly related to each of the variables following partialling. This is exactly what is expected in the case of well-measured, homogeneous constructs, and this case poses little interpretive difficulty because all of the relations present after partialling were present before.

Interpretation becomes more difficult when some relations between the independent variable (IV) and the constructs comprising its nomological network are reduced following partialling but others remain unchanged. Imagine that, before partialling, X_1 is moderately correlated with measures of excitement seeking, aggression, and empathy but, following partialling of X_2, X_1 is only related to empathy. To say in this case that X_1 is a measure of empathy would be incomplete because unpartialled X_1 measures excitement-seeking and aggression as well as empathy. This problem is not insurmountable and statements can be made about incremental prediction and unique variance, but it is more difficult to interpret the residual in terms of the original construct because the residual represents only certain features of the original construct.

The most difficult interpretive problem, however, occurs in the face of suppression when relations between an independent variable and a dependent variable increase or change direction following partialling (e.g., Cohen & Cohen, 1983; Darlington, 1968; Horst, 1941). This situation poses the greatest interpretive hazard because a relation that did not exist or did not exist as strongly before partialling is now uncovered and this relation, therefore, cannot be attributed to the original construct.

Suppression is not always problematic and often has been studied and used explicitly in the areas of educational and occupational testing and, more recently, in the area of personality research (Paulhus, Robins, Trzesniewski, & Tracy, 2004). Horst (1941) described the classic example of suppression in using measures of mechanical and verbal ability to predict pilot performance. When verbal ability was added to the regression of pilot performance on mechanical ability, the relation between mechanical ability and pilot performance increased. This occurred because the test of mechanical ability required verbal skills to read the directions and the extraneous verbal ability variance in mechanical ability was suppressing the relation with pilot

performance. Although this is an example in which suppression does not preclude substantive interpretation, several characteristics of this case are important to note. First, this is a case in which the sole purpose is prediction. The researchers are not trying to say anything about the underlying mechanical ability construct. Second, the relation to pilot performance cannot be ascribed to the unpartialled test of mechanical ability. Third, adequate interpretation of these findings requires knowledge of the content of the individual measures as well as their overlap.

Examples

Although the discussion to this point may suffice to make the general point that researchers must be wary in their interpretation of results from analyses in which one variable is partialled from another, the issues are perhaps best understood concretely as they arise in particular research contexts. In the remainder of the article, we present a number of examples to illustrate the perils of partialling. The examples represent the analytic situation in which these perils are likely to be highest–when highly correlated scales from the same instrument are partialled from one another. Although examples abound, we have chosen three from the literature on aggression and psychopathy.

Reactive-proactive aggression. In 1987, Dodge and Coie proposed a distinction between reactive and proactive aggression. Reactive aggression refers to hostile acts displayed in response to a perceived threat or provocation; proactive aggression refers to an aversive act that is conducted without provocation and with instrumental intent. To measure these constructs, they developed a teacher-rating scale that described reactively and proactively aggressive behaviors and that has been used in a number of studies. Proactive aggression has been argued to be specifically related to positive outcome expectancies for violence (Smithmyer, Hubbard, & Simons, 2000), greater delinquency-related violence (Brendgen, Vitaro, Tremblay, & Lavoie, 2001), greater self-efficacy (Dodge, Lochman, Harnish, Bates, & Pettit, 1997), and fewer internalizing problems (Dodge et al., 1997). In addition, it has been argued that reactive aggression is specifically related to peer rejection (Coie, Dodge, Terry, & Wright, 1991), a hostile attributional bias (Dodge, Price, Bachorowski, & Newman, 1990), and dating-related violence (Brendgen et al., 2001). Many of these conclusions, however, are based primarily on the results from partial correlation analyses. Although the scales possess adequate reliability (values of coefficient alpha between .80 and .90), they are highly correlated (correlations between .60 and .80), which sets the stage for considerable discrepancy between the original variable and the variable that results from partialling. Such discrepancy is evident in articles that provide both the zero-order and partial correlations (e.g., Brendgen et al., 2001; Poulin & Boivin, 2000; Smithmyer et al., 2000).

Childhood psychopathy. Frick and Hare (2001) developed the Antisocial Process Screening Device (APSD) as a childhood extension of the Psychopathy Checklist–Revised (PCL-R; Hare, 2003), which has been widely used to measure psychopathic characteristics in adults. In 1994, Frick, O'Brien, Wootton, and McBurnett suggested that the APSD is underlaid by two factors. One factor, labeled Callous–Unemotional, was argued to index the features traditionally associated with psychopathy in adults, including lack of guilt and shallow emotions. The second factor, labeled Impulsivity/Conduct Problems, was argued to index overt behavioral characteristics such as impulsivity, poor impulse control, and delinquent behavior. Although recent work has suggested that both three- and four-factor models can be adequately fit to the data, we rely on the two-factor model because most of the previous validation work has been done using this model. It has been argued that the Impulsivity/Conduct Problems dimension is specifically related to conduct problems, disruptive behavior disorders, and high levels of anxiety, whereas it has been argued that the Callous–Unemotional dimension is specifically related to a reward-dominant response style (O'Brien & Frick, 1996), low levels of anxiety (Frick et al., 1994; Frick, Lilienfeld, Ellis, Loney, & Silverthorn, 1999), high levels of thrill- and adventure-seeking (Frick et al., 1994), and fearlessness (Frick et al., 1999). As with reactive and proactive aggression, many of these conclusions are based on results from partial correlation analyses (e.g., Frick et al., 1994, 1999; Loney, Frick, Clements, Ellis, & Kerlin, 2003). Given that coefficient alphas for these scales range from .70 to .80 and the correlations between the two factor scales range from .50 to .60, the likelihood of significant discrepancies between the zero-order and partialled relations is high. In fact, several studies have noted explicitly the existence of cooperative suppressor effects (cf. Frick et al., 1999; Loney et al., 2003; Pardini, Lochman, & Frick, 2003).

PCL-R. The PCL-R (Hare, 1991) is a symptom construct rating scale designed to assess psychopathy among criminal offenders and forensic patients. Although three- and four-factor models are gaining ascendance (see Cooke & Michie, 2001; Hare, 2003), much previous work has used a two-factor model for the PCL-R (Harpur, Hakstian, & Hare, 1988). The first factor is described as "a constellation of personality traits that many clinicians consider the core of psychopathy," whereas the second factor is described as being composed of items that "describe a chronically unstable and antisocial lifestyle beginning at an early age" (Harpur et al., 1988, p. 745). Research aimed

at identifying the differential correlates of each factor has shown that Factor 1 is more strongly related than Factor 2 to global clinical ratings of psychopathy, dominance, and low anxiety. On the other hand, compared to Factor 1, Factor 2 is more strongly related to antisocial personality disorder ratings, low socialization, psychoticism, disinhibition, low social class, poor educational achievement, worse institutional behavior, and higher likelihood of recidivism (Harpur, Hare, & Hakstian, 1989). Although this previous work has relied primarily on the differences between zero-order correlations as the means of identifying differential correlates, some research has resorted to partial correlation analyses (e.g., Patrick, 1995; Verona, Patrick, & Joiner, 2001). Given the reliabilities of the factor scales (around .80) and the relatively high correlation between them (from .50 to .60 in most studies), the likelihood of significant discrepancy is again high.

Illustrative Analyses

The difficulty of interpreting results from multiple regression and partial correlation analysis in terms of the original constructs hinges largely on how similar the partial relations are to the zero-order relations. In the remainder of the article, we document empirically the high likelihood of discrepancy between zero-order and partial relations and illustrate more completely the possible interpretive problems that arise as a result of this discrepancy. Specifically, we present findings from analyses in which we compare zero-order and partial relations involving the constructs just reviewed from the aggression and psychopathy literatures. For each set of constructs we first compare zero-order and partial relations among the items assessing a single given construct (e.g., reactive aggression). We then compare zero-order and partial relations of the constructs with a set of outcomes that constitute a small nomological network. In doing so, we document the possibility of a high degree of discrepancy between zero-order and partial correlations.

The nomological network with which we surround the above constructs is drawn primarily from the personality domain. We use the four dimensions from the Eysenck Personality Questionnaire (EPQ; i.e., Extraversion, Neuroticism, Psychoticism, and the Lie Scale), the eight subscales from Lilienfeld's Psychopathic Personality Inventory (PPI; i.e., Egocentricity, Social Potency, Fearlessness, Coldheartedness, Impulsive Noncomformity, Blame Externalization, Nonplanfulness, and Stress Immunity), and the psychopathy and aggression scales. Although the PPI is a well-validated measure of psychopathy, its subscales represent traits that are found in many general models of personality functioning, allowing them to serve as elements of the nomological network.

Based on previous work examining reactive/proactive aggression, the APSD, and the PCL-R, some general, but tentative, predictions can be made; these predictions are qualified because previous research has, at times, relied heavily on partialling to draw conclusions. In terms of aggression, reactive aggression should be more strongly correlated than proactive aggression with measures of negative affect and impulsivity; in contrast, proactive aggression should be more strongly correlated with psychopathy. For the two psychopathy measures, Factor 1, "callous/unemotional" for the APSD and "selfish, callous, remorseless use of others" for the PCL-R, should be related to traits that assess a lack of concern for others, egocentricity, and low anxiety and to proactive rather than reactive aggression. Factor 2, "impulsive/conduct problems" for the APSD and "antisocial/deviant lifestyle" for the PCL-R, should be most strongly related to measures of impulsivity, deviance, and high negative affect and to reactive rather than proactive aggression.

METHOD

Participants

Participants were 696 men incarcerated at a minimum security prison. Fifty-seven percent were African American and the remainder were Caucasian; the average age was 29.1 years. Data from some participants were used in published research reports (passive avoidance learning, Newman & Schmitt, 1998; emotion processing, Lorenz & Newman, 2002; conflict monitoring, Newman, Schmitt, & Voss, 1997; correlational studies involving the correspondence among separate measures of psychopathy, Brinkley, Schmitt, Smith, & Newman, 2001). Potential participants were excluded if they were diagnosed as psychotic or having a bipolar disorder, were using prescribed psychotropic medication, or had performed below the fourth-grade level on the prison's standardized measures of reading and math achievement. Informed consent was obtained both orally and in written form. Participants were informed that their decisions regarding participation would not become part of their record or affect their status in the institution.

Procedure

Individuals agreeing to participate were interviewed and, following a review of their file, rated on the PCL-R (Hare, 2003). All participants also completed the PPI (Lilienfeld & Andrews, 1996) and the EPQ (Eysenck & Eysenck, 1994). In addition, 364 of the inmates completed the APSD (Frick & Hare, 2001), whereas the

remaining 332 completed the Proactive-Reactive Aggression Questionnaire (PRAQ; Raine et al., in press). Paper-and-pencil measures were administered across several days to avoid fatigue.

Measures

PCL-R. The PCL-R (Hare, 2003) is a 20-item symptom construct rating scale completed by interviewers following a semistructured interview and file review. Each item is scored from 0 (*item does not apply*) to 2 (*item applies*) depending on how well the description of the item matches the behavior and personality of the individual. The PCL-R yields three scores: total, Factor 1, and Factor 2; the two subscales are employed in the present study. Although there is disagreement as to what the two factors assess (see Lynam, 2002), items comprising Factor 1 are often referred to as interpersonal and affective characteristics, whereas those assessing Factor 2 are often referred to as impulsive, antisocial, and deviant behavior. Previous studies using these data have consistently reported interrater reliabilities better than .70. In the present sample, coefficient alphas for the two scales were .77 and .61 for Factors 1 and 2, respectively; the correlation between the two scales was .50. The PCL-R is the current gold standard for the assessment of psychopathy. It has excellent psychometric properties, is strongly related to offending, shows incremental predictive utility in relation to recidivism, and has been shown to relate to a variety of underlying emotional, cognitive, and behavioral processes (see Hare, 2003).

APSD. The APSD (Frick & Hare, 2001) is a 20-item behavior rating scale designed to be a childhood extension of the PCL-R. Each item on the APSD is scored 0 (*not at all true*), 1 (*sometimes true*), or 2 (*definitely true*). The APSD yields two subscales, the first subscale (6 items) has been called Callous-Unemotional, whereas the second (10 items) has been referred to as Impulsivity/Conduct Problems. Scores on the APSD have shown good psychometric properties, convergence with other psychopathy measures, positive correlations with measures of antisocial behavior, and expected relations to a number of putative underlying psychopathic processes (for a review, see Lynam & Gudonis, 2005).

Although the APSD was originally designed for use with parents and teachers of children and adolescents, we include it in the present study because previous research has relied heavily on the partialling process to draw conclusions about the scale. In addition, our use of the APSD in an adult sample is consistent with two recent reports. Kruh, Frick, and Clements (2005) used a self-report version of the APSD in a sample of inmates age 16 to 21 years, and Benning, Patrick, Salekin, and Leistico (2005) used it in a sample of undergraduates; both sets of researchers found that the APSD performed as anticipated. We would note, however, the important distinction between prior use and proper use; our use, although defensible in the present context, represents quite an extension the original application. Coefficient alphas for the two subscales were .47 and .64 for Factors 1 and 2, respectively; the correlation between the scales was .50.

PRAQ. The PRAQ (Raine et al., 2006) is a 26-item self-report questionnaire designed to assess proactive and reactive aggression. Items were generated based on the items contained in teacher-rating measures of proactive-reactive aggression (Brown, Atkins, Osborne, & Millnamow, 1996; Dodge & Coie, 1987) and the conceptual and theoretical literature on proactive and reactive aggression. Thirteen items assess proactive aggression, or aggression without immediate provocation (e.g., had fights to show who was on top, hurt others to win a game, and used force to obtain money or things from others). Thirteen items assess reactive aggression or defensive reaction to a threatening situation (e.g., gotten angry when frustrated, gotten angry or mad when you lost a game, and hit others when teased). For each item, participants indicate how often they have engaged in each kind of behavior. Responses range from 0 (*never*) to 5 (*always or almost always*). Coefficient alpha was .89 for each subscale. Consistent with previous work using teacher reports or peer nominations, the correlation between the two scales was .80 (e.g., Price & Dodge, 1989). Scores from this questionnaire have been shown to relate important behavioral and personality correlates including violence, delinquency, psychopathy, impulsivity, and psychopathology (e.g., Raine et al., 2006).

EPQ. The EPQ (Eysenck & Eysenck, 1994) is 90-item, self-report measure designed to assess the basic traits of personality. The EPQ contains four subscales: Extraversion (E), Neuroticism (N), Psychoticism (P), and Lie. Although originally conceived as a validity scale, research demonstrates that the Lie Scale may be better conceived as an index of social conformity (e.g., Eysenck & Eysenck, 1994). Coefficient alphas were .83, .86, .67, and .80 for E, N, P, and Lie, respectively.

PPI. The PPI (Lilienfeld & Andrews, 1996) is a 187-item, self-report measure designed to assess eight of the core elements of psychopathy. For each item, respondents indicate how true a given item is of them; responses range from 1 (*true*) to 4 (*false*). The eight subscales are as follows: Machiavellian Egocentricity assesses narcissistic and ruthless attitudes in interpersonal relations (30 items; e.g., "I always look out for my own interests before worrying about those of the other guy"; $\alpha = .88$). Social Potency assesses the tendency to be charming and adept at influencing others (24 items; e.g., "Even when others are

upset with me, I can usually win them over with my charm"; α = .83). Coldheartedness assesses a propensity toward callousness and unsentimentality (21 items; e.g., "I have had crushes on people that were so intense they were painful," reverse-scored; α = .80). Carefree Nonplanfulness assesses an absence of forethought (20 items; e.g., "I often make the same errors in judgment over and over again"; α = .85). Fearlessness assesses a lack of anticipatory anxiety in the face of potential harm and a willingness to take risks (19 items; e.g., "Making a parachute jump would really frighten me," reverse-scored; α = .84). Blame Externalization assesses the tendency to view others as the source of one's problems and to rationalize one's behavior (18 items; e.g., "I usually feel that people give me the credit that I deserve," reverse-scored; α = .84). Impulsive Nonconformity assesses a lack of concern regarding social mores (17 items; e.g., "I sometimes question authority figures 'just for the hell of it;' " α = .76). Finally, Stress Immunity assesses an absence of reaction to anxiety-provoking events (11 items; e.g., "I can remain calm in situations that would make many other people panic"; α = .74). The PPI also contains three validity scales designed to assess various response sets; these were not used in the present research. The PPI has been shown to converge with other psychopathy assessments, to correlate with reports of antisocial and deviant behavior, and to show expected patterns of correlations with basic dimensions of personality in both noncriminal and criminal samples (for a review, see Lilienfeld & Fowler, 2005).

RESULTS

Item-Level Analyses

One means of evaluating the performance of a measure is to examine the internal consistency of its items. In these analyses, we evaluated the degree to which the internal consistency of items composing the various scales changes before and after partialling. To evaluate this property of the aggression and psychopathy scales, the items contributing to each scale were first residualized. This was accomplished by regressing each item from a given subscale onto its subscale counterpart and saving the residuals. For example, each proactive aggression item was regressed onto the reactive aggression scale and the residual from each analysis was saved. The relations among the residualized items, specifically, the corrected item-total correlations, were then compared to the relations among the raw items via an intraclass correlation (ICC). The ICC was calculated as a double-entry correlation and was employed because it takes into account both similarity in direction and magnitude (Haggard, 1958). Such a comparison provides an index of how similarly the items on a scale relate to each other before and after partialling.

Results indicated substantial dissimilarity in the internal consistency of the scales before and after partialling. For the aggression subscales, ICCs between corrected item-total correlations for raw and residualized items were –.42 and –.21 for proactive and reactive aggression, respectively. For the subscales of the APSD, ICCs were .44 and .78 for Factors 1 and 2, respectively. For Factors 1 and 2 of the PCL-R, ICCs were .46 and .57.

Scale-Level Analyses

In contrast to the item-level analyses, which draw attention to the implications of partialling for the relations among items on a scale, scale-level analyses shed light on the impact of partialling on the relations between a scale and measures of other constructs in its nomological network. For scale-level analyses, we examined the relations of each subscale, before and after partialling the other relevant subscale, to variables comprising the nomological network. The correspondence or similarity between the coefficients (i.e., the nomological networks) from the two analyses (before and after partialling) was directly assessed by computing an ICC on the coefficients for the raw and residualized scales. For example, the correspondence between the nomological networks for the original and residualized proactive aggression scale was calculated by double-entering the first two columns of Table 2 and computing a correlation.[4] This analysis is important because to the extent that the networks of meaning differ across the original and residualized scales, the constructs measured by the scales can be said to differ. The results, presented in Tables 1 through 3, document the fact that the residualized scores often relate differently to other variables than their original score counterparts.

Tables 1 through 3 provide the unstandardized regression coefficients from analyses in which pairs of scales from the PRAQ, APSD, and PCL-R are used to predict scores on scales from the EPQ, PPI, and where available, the other focal measures (e.g., the PCL-R factors for reactive and proactive aggression). Coefficients for the raw scales are taken from analyses in which the given subscale was used as the sole predictor. Coefficients for the residual scales are taken from an analysis in which both subscales were entered. It is possible to compare the change in the coefficients across the raw and residualized scales by examining the difference as a function of the standard error for the third variable effect (see MacKinnon, Krull, & Lockwood, 2000); these results are provided in Tables 1 through 3.[5] It is also possible to compare more holistically the nomological networks of the raw and residualized variables through the use of an

TABLE 1
Relations Between Reactive/Proactive Aggression Scores and Personality Before and After Partialling

Personality Scale	*Raw Proactive*	*Residual Proactive*	z *Value*	*Raw Reactive*	*Residual Reactive*	z *Value*
EPQ-Extraversion	–.023	–.050	< 1	–.005	.031	< 1
EPQ-Neuroticism	**.202**	.038	3.20**	**.212**	**.185**	< 1
EPQ-Psychoticism	**.170**	**.168**	< 1	**.122**	.002	4.44***
EPQ-Lie	**–.249**	–.083	4.48***	**–.247**	**–.188**	1.79
PPI-Egocentricity	**.884**	**.283**	6.98***	**.907**	**.695**	2.26*
PPI-Social potency	.127	.032	< 1	**.134**	.110	< 1
PPI-Fearlessness	**.370**	–.005	4.07***	**.430**	**.434**	< 1
PPI-Coldheartedness	.099	**.342**	2.85**	–.025	**–.280**	3.23**
PPI–Imp. nonconform	**.429**	**.225**	3.16**	**.403**	**.235**	2.80**
PPI–Blame external	**.430**	.168	3.10**	**.429**	**.303**	1.62
PPI-Nonplanfulness	**.300**	**.393**	1.10	**.186**	–.107	3.70***
PPI–Stress immunity	**–.142**	.046	3.93***	**–.184**	**–.219**	< 1
PCL-R Factor 1	**.064**	**.122**	2.03*	.024	**–.068**	3.46***
PCL-R Factor 2	**.190**	**.129**	2.07*	**.167**	**.070**	3.47***
Similarity coefficients						
Resid Proactive Agg.	.406					
Raw Reactive Agg.	.981	.299				
Resid Reactive Agg.	.769	.023		.869		

NOTE: Numbers in top part of table are unstandardized regression coefficients. Coefficients significant at $p < .05$ are indicated in bold. The z value is the test of the difference between the raw and residualized coefficients. EPQ = Eysenck Personality Questionnaire; PPI = Psychopathic Personality Instrument; PCL-R = Psychopathy Checklist–Revised; Resid = Residual; Agg. = Aggression. Similarity coefficients are intraclass correlation coefficients that index the similarity between unstandardized regression coefficients from equations predicting EPQ and PPI subscales and selected other subscales either ignoring the corresponding subscale (Raw) or accounting for overlap with it (Residual). For example, .981 represents the similarity between columns 1 and 4.
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 2
Relations Between APSD Factor 1 and 2 Scores and Personality Before and After Partialling

Personality Scale	*Raw Factor 1*	*Residual Factor 1*	z *Value*	*Raw Factor 2*	*Residual Factor 2*	z *Value*
EPQ-Extraversion	**–1.63**	**–2.46**	1.80	.415	**1.61**	2.64**
EPI-Neuroticism	**2.59**	–.943	5.97***	**6.37**	**6.83**	< 1
EPI-Psychoticism	**4.43**	**3.10**	4.16***	**4.09**	**2.58**	5.07***
EPI-Lie	**–4.57**	**–1.35**	7.23***	**–6.88**	**–6.22**	1.99*
PPI-Egocentricity	**29.53**	**20.81**	6.35***	**26.86**	**16.78**	7.10***
PPI–Social potency	**3.95**	**2.98**	< 1	3.32	1.87	< 1
PPI-Fearlessness	**10.27**	3.26	6.06***	**15.05**	**13.47**	1.68
PPI-Coldheartedness	**8.69**	**10.85**	2.28*	1.12	**–4.16**	5.19***
PPI–Imp. nonconform	**11.63**	**6.67**	5.90***	**12.80**	**9.56**	4.27***
PPI–Blame external	**7.73**	**4.31**	3.65**	**8.68**	**6.59**	2.39*
PPI-Nonplanfulness	**9.10**	**6.00**	3.41***	**8.90**	**5.98**	3.67***
PPI–Stress immunity	**–2.87**	–.247	4.26***	**–5.16**	**–5.05**	< 1
PCL-R Factor 1	**1.74**	.78	2.92**	**2.21**	**1.83**	1.24
PCL-R Factor 2	**3.29**	**1.45**	5.23***	**4.22**	**3.52**	2.30*
Similarity coefficients						
Resid APSD F1	.850					
Raw APSD F2	.936	.681				
Resid APSD F2	.738	.531		.895		

NOTE: Numbers in top part of table are unstandardized regression coefficients. Coefficients significant at $p < .05$ are indicated in bold. The z value is the test of the difference between the raw and residualized coefficients. APSD = Antisocial Process Screening Device; EPQ = Eysenck Personality Questionnaire; PPI = Psychopathic Personality Instrument; PCL-R = Psychopathy Checklist–Revised; Resid = Residual. Similarity coefficients are intraclass correlation coefficients that index the similarity between unstandardized regression coefficients from equations predicting EPQ and PPI subscales and selected other subscales either ignoring the corresponding subscale (Raw) or accounting for overlap with it (Residual).
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 3
Relations Between PCL-R Factor 1 and 2 Scores and Personality Before and After Partialling

Personality Scale	*Raw Factor 1*	*Residual Factor 1*	z *Value*	*Raw Factor 2*	*Residual Factor 2*	z *Value*
EPQ-Extraversion	.063	.027	1.25	.085	.072	< 1
EPQ-Neuroticism	–.044	**–.186**	3.97***	**.191**	**.285**	2.65**
EPQ-Psychoticism	**.149**	.030	5.64**	**.253**	**.238**	< 1
EPQ-Lie	**–.116**	.084	6.90***	**–.359**	**–.402**	1.66
PPI-Egocentricity	**.832**	.202	6.52***	**1.36**	**1.26**	1.16
PPI–Social potency	**.599**	**.473**	1.92	**.490**	.250	3.54***
PPI-Fearlessness	.188	**–.294**	6.64***	**.812**	**.961**	2.25*
PPI-Coldheartedness	**.359**	**.289**	1.17	**.286**	.139	2.42*
PPI–Imp. nonconform	**.419**	.024	7.92***	**.870**	**.882**	< 1
PPI–Blame external	**.382**	.113	4.32***	**.593**	**.536**	< 1
PPI-Nonplanfulness	.140	–.198	5.42***	**.571**	**.672**	1.70
PPI–Stress immunity	**.149**	**.251**	2.81**	–.075	**–.203**	3.5***
Proactive Agg.	**.424**	–.061	5.87***	**.974**	**.997**	< 1
Reactive Agg.	.185	**–.364**	6.02***	**.990**	**1.13**	2.29*
PSD Factor 1	**.016**	.002	4.78***	**.028**	**.027**	< 1
PSD Factor 2	**.024**	.005	5.96***	**.039**	**.036**	< 1
Similarity coefficients						
Resid PCL-R F1	.201					
Raw PCL-R F2	.509	–.388				
Resid PCL-R F2	.399	–.465		.976		

NOTE: Numbers in top part of table are unstandardized regression coefficients. Coefficients significant at $p < .05$ are indicated in bold. The z value is the test of the difference between the raw and residualized coefficients. PCL-R = Psychopathy Checklist–Revised; EPQ = Eysenck Personality Questionnaire; PPI = Psychopathic Personality Instrument; PSD = Psychopathy Screening Device; Agg. = Aggression; Resid = Residual. Similarity coefficients are intraclass correlation coefficients that index the similarity between unstandardized regression coefficients from equations predicting EPQ and PPI subscales and selected other subscales either ignoring the corresponding subscale (Raw) or accounting for overlap with it (Residual).
*$p < .05$. **$p < .01$. ***$p < .001$.

ICC; these results are presented at the bottom of the tables. As noted earlier, the ICC is particularly revealing in this case because it takes into account not only similarity in profile shape but also similarity in magnitude.

Proactive-reactive aggression. Table 1 provides the results for the aggression subscales. The ICC coefficients at the bottom index the extent to which proactive and reactive aggression scores relate similarly to the EPQ and PPI scales before and after partialling. First, note that the value of .981 in the first column of similarity inidices indicates that, before partialling, the nomological network as we have defined it is virtually identical for the proactive and reactive scales. By comparison, the last value in the second column, .023, indicates no overlap at all in the nomological networks of the two when each is partialled from the other. This dramatic difference in the nomological network of the raw and residualized scores is attributable to the fact that the raw and residualized scores for each scale clearly are not the same; the ICC for the nomological networks of raw and residualized scores is .406 and .869 for proactive and reactive aggression, respectively. In fact, raw reactive and proactive aggression are more similar to each other in their nomological networks than either is to its own residualized counterpart.

The discrepancy between the raw and residualized scores is underscored through an examination of Table 1. For proactive aggression, 10 of the 14 coefficients change significantly following the partialling process. Most of these represent decreases, often fairly large ones, in the size of the coefficients; for example, the coefficients for Neuroticism, Lie Scale, Egocentricity, Impulsive Nonconformity, and Blame Externalization change by at least 3 standard errors following partialling. For other scales, and more problematically, there is a significant increase in the size of the relations following partialling; this is true for PCL-R Factor 1 and Coldheartedness, for which the zero-order coefficient was not significantly different from zero. For two scales, Fearlessness and Stress Immunity, there is a change in the direction of the relations following partialling. Although the nomological networks for the reactive subscale are more similar to one another, 7 of 14 coefficients undergo significant change following partialling. Four represent significant decreases (Psychoticism, Egocentricity, Impulsive Nonconformity, and PCL-R Factor 2), one represents a significant increase (Coldheartedness), and two represent changes in direction (Nonplanfulness and PCL-R Factor 1). The raw and residualized scales are measuring different constructs.

APSD Factors 1 and 2. ICC coefficients presented at the bottom of Table 2 reveal that the partialled subscales are not equivalent to the raw subscales. Again, at the raw score level, there is very little difference in the nomological networks of the raw scores representing Factor 1 and Factor 2; the ICC of .936 is extremely high. It also can be seen that the partialling process has yielded two residualized variables with more differentiated nomological networks (see the ICC of .531 at the bottom of column 2). Although the difference is not quite as dramatic for the APSD as for the PRAQ above, the nomological networks of the raw scores are again more similar to each other (ICC = .936) than the networks of either of the raw scores is to its residualized counterpart (ICCs = .85 and .90 for Factors 1 and 2, respectively).

Table 2 reveals the nature of the differences. For Factor 1, 12 of 14 coefficients change significantly following partialling. Ten of these represent decreases in the size of the relations; seven of these, Psychoticism, Lie Scale, Egocentricity, Fearlessness, Impulsive Nonconformity, Stress Immunity, and PCL-R Factor 2, represent decreases of more than 4 standard errors. For Neuroticism, there was a reversal in sign. The size of the relation increased in the case of Coldheartedness. In the case of Factor 2, 9 of the 14 coefficients underwent significant changes following the partialling process. Seven of the changes represent decreases in the size of the relations, with three of these being greater than 4 standard errors: Psychoticism, Egocentricity, and Impulsive Noncomformity. The changes for both Extraversion and Coldheartedness represent increases; in fact, the coefficients become significantly different from zero only after partialling. Again, the raw and residualized scales are measuring different constructs.

PCL-R Factors 1 and 2. Coefficients in the bottom panel of Table 3 again reveal that the partialling process changes the nature of the variables under study. In the case of the PCL-R, compared to previous variables, there is more divergence in the nomological networks surrounding the raw scales (see ICC of .50 in Column 1), but the partialling process has again yielded two residualized variables with even more differentiated nomological networks (see ICC of –.47). Finally, the ICCs between the coefficients for the raw and residualized variables indicate that although the scores behave similarly for Factor 2 (.976), there is very little similarity in the nomological networks or raw and residualized Factor 1 (.201).

Table 3 provides detail regarding the differences. In the case of Factor 1, 13 of 16 coefficients are significantly changed following partialling. Six of these changes, Psychoticism, Egocentricity, Impulsive Nonconformity, Blame externalization, and APSD Factors 1 and 2, represent decreases; in all of these cases, significant coefficients became nonsignificant following partialling and all decreases were greater than 4 standard errors. For Neuroticism and Stress Immunity, there were significant increases in the sizes of the relations; in the case of Neuroticism, the relation was only significant following partialling. For the remaining five variables, Lie, Fearlessness, Nonplanfulness, and both forms of aggression, there was a reversal in sign. For Fearlessness and Reactive Aggression, nonsignificant positive relations became significant negative relations following partialling. In the case of Factor 2, only 6 of the 16 coefficients were significantly different following partialling. Of interest, four of the six significant changes represent significant increases. Certainly in the case of Factor 1, the raw and residualized scales are measuring different constructs.

DISCUSSION

In this article, we sought to demonstrate the interpretive difficulties posed by partial regression analyses in the case of two highly correlated predictors. We argued that the main difficulty lies in interpreting the results from the partial analyses in terms of the original scales. Using the subscales from three widely used instruments, we showed that the nomological networks surrounding raw and residualized scores often bore little resemblance to one another. The comparison of relations between scales before and after partialling was particularly revealing, showing that some relations disappeared, others remained the same, and still others emerged only after partialling. Although these findings do not necessarily argue against partialling when two highly correlated scales are being compared, they do highlight the logical and interpretive problems that arise.

Each of the three scales was a likely candidate for high levels of discrepancy between raw and residualized scores. Interscale correlations ranged from .50 to .80, whereas coefficient alphas ranged from .47 to .89. Although most of the alphas were consistent with previous studies, the alpha for PCL Factor 2 was somewhat lower than previously observed intensifying concern regarding interpretation of the partials . The degree of discrepancy was relatively high as shown by a comparison of correlations derived from zero-order and partial analyses. In terms of internal consistency (i.e., item-total correlations for raw and residualized scales), the highest observed similarity was .78 for APSD Factor 2, whereas the lowest was –.42 for proactive aggression. There also was considerable discrepancy in terms of the nomological network surrounding the raw and residualized scales; although the highest similarity was quite high, .98 for PCL-R Factor 2, and the lowest was quite low, .20 for PCL-R Factor 1.

It is tempting and possible in each case to pull from the zero-order and partial analyses results that are consistent with prior research and theory. For example, it is possible to interpret results for PCL-R Factor 1 as follows: PCL-R Factor 1 is associated with low levels of Neuroticism, Reactive Aggression, and social conformity; high levels of Psychoticism, Egocentricity, Social Potency, Coldheartedness, Impulsive Noncomformity, Blame Externalization, Stress Immunity, and Proactive Aggression; and a tendency toward Fearlessness. All of these relations are consistent with the theoretical framework surrounding psychopathy; unfortunately, such an interpretation is invalid, blending as it does results from the zero-order and partial analyses. Only the residualized PCL-R Factor 1 scores are negatively related to Neuroticism and Reactive Aggression. Similarly, the relations to Psychoticism, Social Conformity, Egocentricity, Impulsive Nonconformity, Blame Externalization, and Proactive Aggression are present only in relation to the unresidualized scale. Finally, although there is a tendency for the unresidualized PCL-R Factor 1 scores to be related to Fearlessness, the residualized scales are actually significantly negatively related to Fearlessness, which runs contrary to many theories (Lykken, 1995).

A more conservative interpretation of the results would begin first by noting the similarity between the relations for Factor 1 and Factor 2 at the zero-order level (see similarity coefficient of .509 in Table 1). At this level, both factors are significantly positively related to Psychoticism, Egocentricity, Social Potency, Coldheartedness, Impulsive Nonconformity, Blame Externalization, Proactive Aggression, and both APSD factors; both factors are significantly negatively related to the Lie Scale. Both factors seem to assess an antagonistic, dominant, and egocentric interpersonal orientation interpersonal style, as well as some degree of impulsivity and unconventionality. In addition, Factor 2 is also positively related to Neuroticism, Fearlessness, Nonplanfulness, and Reactive Aggression. At the zero-order level, Factor 1 has no unique correlates. At this level, the factors are distinguished primarily by the unique relations that Factor 2 bears to traits associated with emotional reactivity and poor impulse control.

The second step involves discussion of the partial coefficients. At this point, the similarity or dissimilarity between raw and residualized relations should be noted. On one hand, the relations for the raw and residualized Factor 2 scores are quite consistent with one another (see .976 similarity coefficient in Table 1). After removing the variance shared with Factor 1, Factor 2 remains significantly positively related to Neuroticism, Psychoticism, Egocentricity, Fearlessness, Impulsive Nonconformity, Blame Externalization, Nonplanfulness, Proactive and Reactive Aggression, and both scales of the APSD; in addition,, it remains negatively related to the Lie Scale. On the other hand, the relations for raw and residualized Factor 1 scores are quite different from one another (see .201 similarity in Table 1); in fact, the nomological network for raw Factor 1 is more similar to the networks for raw and residual Factor 2 than to the network for residual Factor 1. Factor 1 scores remain significantly related only to Social Potency, Coldheartedness, and Stress Immunity.

Most important, particularly for Factor 1, the partialling process has exposed several relations that did not exist at the zero-order level. Residualized Factor 1 scores are significantly negatively related to Neuroticism, Fearlessness, Nonplanfulness, and Reactive Aggression. All of these relations are opposite to those observed for the raw scale; they suggest that the residual scores, in contrast to the raw scores, assess good impulse control and low neuroticism.

The implications drawn from these two interpretations are divergent. The first interpretation, mixing raw and residualized relations, ignores the extremely high degree of overlap between the scales in their relations to the criteria; this problem is even greater for the subscales of APSD and PRAQ, which have virtually identical nomological nets at the raw score level. In addition, as noted earlier, the first interpretation blends results from the raw and residualized scales, highlighting results consistent with theory and ignoring results inconsistent with theory. Such blending is hard to justify on methodological or statistical grounds. For the PCL-R, what is the justification for reporting the negative relation between residualized F1 and neuroticism as supporting the theory while at the same time ignoring the negative relation between residualized F1 and Fearlessness? In the case of the PRAQ, the positive relations between raw proactive aggression and Fearlessness and residualized proactive aggression and Coldheartedness are both consistent with theory. However, these relations are not present at the same time; the relation to Fearlessness is only present for the residualized scale, whereas the relation to Coldheartedness is only present for the raw scale. On what methodological grounds can one argue that both of these relations characterize the same scale? It would seem as legitimate to argue that proactive aggression is not related to either of these criteria because there is no relation between the raw scale and Fearlessness or the residualized scale and Coldheartedness. The second, more conservative interpretation preserves the distinctions between the raw and residualized scales.

What Is to Be Done?

First, we offer, homiletically, that researchers employ differentiable, reliable, homogeneous, and well-understood constructs and measures. These are conditions under which partial regression analyses are most informative.

As noted in the introduction, the problems surrounding the substantive interpretation are greatest in the face of highly correlated, heterogeneous, unreliable measures. When these conditions are absent, the problem of substantive interpretation is greatly lessened; in fact, interpretation of the partial relations is relatively straightforward in these cases. It must be recognized that some constructs, including those employed in the present article, are inherently complex and high internal consistency for a total score may not be a desideratum. By narrowing the scope of a variable, one may reduce its validity. In these cases, we argue for creation and examination of more internally consistent subscales; this seems the best approach to reducing or understanding what the partialling process is doing to variables. If this approach is not feasible, interpretive problems can be overcome to some degree through theory (Hoyle & Robinson, 2004), even under undesirable conditions of low reliability, low internal consistency, and high overlap between variables.

Second, under these undesirable conditions, when there is a significant discrepancy between the raw and residual variables, we suggest a decision must be made as to which construct, the original or residualized, the conclusions are meant to apply. This decision determines whether the zero-order or partial correlations are to be examined. If the results are meant to apply to the original scale, then the zero-order correlations are the relevant relations. This would seem to be the case most often because those are the scales that comprise the items as written. If, however, the residualized scales are the targets of conclusions, then interpretation becomes more difficult. For example, it is easy to say that after removing the variance associated with PCL-R Factor 2, PCL-R Factor 1 is associated with low Neuroticism, Social Potency, low Fearlessness (or fearfulness), Coldheartedness, Stress Immunity, and low Reactive Aggression. However, it is much more difficult to say what PCL-R Factor 1 is with the variance associated with PCL-R Factor 2 removed. One cannot look at the original items that make up PCL-R Factor 1; after all, these items define the original scale and not the residualized scale. We also note that aiming the conclusions at the residualized scales greatly complicates the process of validation and reduces the utility of the scale. Because the residualized scale exists only after partialling, all validation work must include both scales and only the partial correlations are relevant; in addition, results are specific to the dependent variable examined and the particular independent variable that is partialled. Moreover, the scale becomes especially difficult, if not impossible, to use in clinical settings where only the raw scores are available. This renders these scales much less useful.

Our third suggestion highlights the necessity of good theory (see Meehl, 1975; Paulhus et al., 2004). To interpret the residualized relations, one must look to theory, which requires an understanding of what each of the original scales represents as well as an understanding of what they share. In the face of suppression, additional theory is necessary to explain why removing variance shared across subscales increases the correlation with the outcome. It is important to note the form that this theory must take because previous investigators have occasionally confused suppression with interaction (e.g., Loney et al., 2003), which can both be conceived as third variable effects. Suppression involves a change in the relation between variables as a function of the presence or absence in the model of a third variable. Interaction involves a change in the relation between variables as a function of the level of the third variable. Thus, an account of suppression must explain why the suppressed relation is not present at the zero-order level and how the inclusion of a third variable frees up variance in the variable of interest.

In the case of the relations for the PCL-R (and APSD), it might be possible to look to the five-factor model (FFM) account of psychopathy for guidance (Lynam, 2002; Lynam & Derefinko, 2005; Lynam et al., 2005). Lynam and colleagues have argued that psychopathy can be understood from the perspective of the FFM, which consists of five broad domains each underlaid by six facets or subtraits (Costa & McCrae, 1992). They have argued that this understanding of psychopathy brings clarity to several issues in the field, including the two-factor structure of the PCL-R. These authors have suggested that both factors assess extremely low Agreeableness but are differentiated by the stronger relations of Factor 2 to low Conscientiousness and high Neuroticism. This model also may offer an account of the suppressor relations observed for PCL-R (and APSD) Factor 1. Interesting divergences within the facets of Neuroticism in their relation to psychopathy have appeared across multiple studies. Psychopathy is positively related to some facets of Neuroticism (e.g., angry hostility and impulsiveness) but negatively related to others (e.g., anxiousness, self-consciousness, and vulnerability). Lynam and colleagues (see especially Lynam et al., 2005) have argued that whereas Factor 2 is associated primarily with high Neuroticism, Factor 1 may contain representations of both poles of Neuroticism. From this perspective, the absence of a zero-order relation between PCL-R Factor 1 and indices of Neuroticism is due to the fact that Factor 1 contains both low and high aspects of Neuroticism, which tend to cancel each other out. When Factor 2 scores, which include only representations of high Neuroticism, are partialled from Factor 1 scores, then the aspects of Factor 1 related to low Neuroticism are freed up.

The fourth recommendation concerns best reporting procedures. Zero-order coefficients should always be

presented in addition to the partial coefficients. This allows reviewers and readers to determine the degree of discrepancy between coefficients produced with the original scales and those produced by partialling. In addition, relations indicative of suppression should always be highlighted. Conclusions drawn from partial analyses must always be couched in terms of the residual variable and not in terms of the original variable.

Finally, we note that other potential statistical approaches might be employed, notably, structural equation modeling. Within such a framework, multiple models can be specified and the changes that occur across these models can be examined. There are advantages to models such as these: they illustrate in a very concrete way what the partialling process is doing and allow removal of error variance at the item level. However, there are disadvantages and limits as well: such approaches require the estimation of multiple models to obtain residual and raw relations, the approaches do not eliminate suppression, and it is difficult to test the raw and residual relations against one another. The present data were subjected to such analyses and results were consistent with those presented.[6]

We wish to underscore that we are not arguing against the use of partialling in all cases. As we noted at the beginning of this article, partialling in the context of multiple regression allows us to test some of our most interesting theories and hypotheses. In many cases, particularly those involving internally consistent variables that are weakly or moderately correlated and do not evince suppressor relations, the interpretations are relatively straightforward. Even suppression is not a necessary contraindication to partialling. As indicated by Paulhus et al. (2004), suppression can be a useful finding when there is theory available to explain the phenomenon. Our argument is aimed primarily at the failure to give sufficient attention to what is left in a variable once the variance it shares with another variable, particularly a highly correlated other variable, is removed.

NOTES

1. This is seen in the formula for the standard error of the unstandardized regression coefficient: $SE_{Bi} = \sqrt{(1 - R_Y^2)/(\mathrm{n} - \mathrm{k} - 1)}\sqrt{1/(1 - R_i^2)}$, where R_Y^2 is the proportion of variance accounted for in the dependent variable by all independent variables and R_i^2 is the proportion of variance accounted for in the independent variable of interest by all other independent variables (i.e., multicollinearity).

2. This correlation is found by the formula $r_{rawres} = \sqrt{1 - r_{raw1raw2}^2}$, where r_{rawres} is the correlation between the raw and residualized scores and $r_{raw1raw2}$ is the correlation between the raw scores.

3. The relation between the reliability of the residualized scores and the reliability of the raw scores is evident in the formula for reliability of the residualized scores, $r_{res1res1} = \dfrac{r_{raw1raw1} - r_{raw1raw2}^2}{1 - r_{raw1raw2}^2}$, where $r_{res1res1}$ is reliability of the residualized scores, $r_{raw1raw1}$ is reliability of the raw scores, and $r_{raw1raw2}$ is the correlation between the two sets of scores. It must be noted that this formula will be inaccurate in cases where systematic error variance is shared across variables.

4. To be more specific, the 14 coefficients in column 1 of Table 3 were appended to column 2, and the coefficients in column 2 were appended to column 1, rendering a data matrix with two columns and 28 rows.

5. This is obtained by taking the square root of the first-order Taylor series (see Sobel, 1982), $\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2$. These coefficients are obtained from the two equations: $Y = \beta_0 + \tau' X + \beta Z + \varepsilon$ and $Z = \beta_0 + \alpha X + \varepsilon$, in which Y is the outcome variable, X is the subscale of interest, and Z is the subscale to be partialled from X.

6. Details of the analysis and results are available from the first author.

REFERENCES

Benning, S. D., Patrick, C. J., Salekin, R. T., & Leistico, A. M. (2005). Convergent and discriminant validity of psychopathy factors assessed via self-report: A comparison of three instruments. *Assessment*, *12*, 270-289.

Brendgen, M., Vitaro, F., Tremblay, R. E., & Lavoie, F. (2001). Reactive and proactive aggression: Predictions to physical violence in different contexts and moderating effects of parental monitoring and caregiving behavior. *Journal of Abnormal Child Psychology*, *29*, 293-304.

Brinkley, C. A., Schmitt, W. A., Smith, S. S., & Newman, J. P. (2001). Construct validation of a self-report psychopathy scale: Does Levenson's SRPS measure the same construct as Hare's PCL-R? *Personality and Individual Differences*, *31*, 1021-1038.

Brown, K., Atkins, M. S., Osborne, M. L., & Milnamow, M. (1996). A revised teacher rating scale for reactive and proactive aggression. *Journal of Abnormal Child Psychology*, *24*, 473-480.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficits. *Psychological Bulletin*, *79*, 380-385.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Coie, J. D., Dodge, K. A., Terry, R., & Wright, V. (1991). The role of aggression in peer relations: An analysis of aggression episodes in boy's play groups. *Child Development*, *62*, 812-826.

Cooke, D. J., & Michie, C. (2001). Refining the construct of psychopath: Towards a hierarchical model. *Psychological Assessment*, *13*, 171-188.

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*, 343-359.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*, 161-182.

Dodge, K. A., & Coie, J. D. (1987). Social information-processing factors in reactive and proactive aggression in children's playgroups. *Journal of Personality and Social Psychology*, *53*, 1146-1158.

Dodge, K. A., Lochman, J. E., Harnish, J. D., Bates, J. E., & Pettit, G. S. (1997). Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth. *Journal of Abnormal Psychology*, *106*, 37-51.

Dodge, K. A., Price, J. M., Bachorowski, J. A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology*, *99*, 385-392.

Eysenck, H. J., & Eysenck, S. B. G. (1994). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: Educational and Industrial Testing Service.

Frick, P. J., & Hare, R. D. (2001). *The Psychopathy Screening Device (PSD)*. Toronto, Canada: Multi-Health Systems.

Frick, P. J., Lilienfeld, S. O., Ellis, M., Loney, B., & Silverthorn, P. (1999). The association between anxiety and psychopathy dimensions in children. *Journal of Abnormal Child Psychology*, *27*, 383-392.

Frick, P. J., O'Brien, B. S., Wootton, J. M., & McBurnett, K. (1994). Psychopathy and conduct problems in children. *Journal of Abnormal Psychology*, *103*, 700-707.

Gordon, R. A. (1968). Issues in multiple regression. *American Journal of Sociology*, *73*, 592-616.

Haggard, E. A. (1958). *Intraclass correlations and the analysis of variance*. New York: Dryden.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised*. Canada: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised* (2nd ed.). Canada: Multi-Health Systems.

Harpur, T. J., Hakstian, A. R., & Hare, R. D. (1988). Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology*, *56*, 741-747.

Harpur, T. J., Hare, R. D., & Hakstian, A. R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment*, *1*, 6-17.

Horst, P. (1941). The role of predictor variables which are independent of the criterion. *Social Science Research Council Bulletin*, *48*, 431-436.

Hoyle, R. H., & Robinson, J. I. (2004). A reasoned approach to measurement in social psychological research. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology*. Thousand Oaks, CA: Sage.

Kruh, I. P., Frick, P. J., & Clements, C. B. (2005). Historical and personality correlates to the violence patterns of juveniles tried as adults. *Criminal Justice and Behavior*, *32*, 69-96.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *Journal of Personality Assessment*, *66*, 488-524.

Lilienfeld, S. O., & Fowler, K. A. (2005). The self-report assessment of psychopathy: Problems, pitfalls, and promises. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 107-132). New York: Guilford.

Loney, B. R., Frick, P. J., Clements, C. B., Ellis, M. L., & Kerlin, K. (2003). Callous-unemotional traits, impulsivity, and emotional processing in adolescents with antisocial behavior problems. *Journal of Clinical Child and Adolescent Psychology*, *32*, 66-80.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304-305.

Lorenz, A. R., & Newman, J. P. (2002). Deficient response modulation and emotion processing in low-anxious Caucasian psychopathic offenders: Results from a lexical decision task. *Emotion*, *2*, 91-104.

Lykken, D. T. (1995). *The antisocial personalities*. Mahwah, NJ: Lawrence Erlbaum.

Lynam, D. R. (2002). Psychopathy from the perspective of the five factor model. In P. T. Costa & T. A. Widiger (Eds.), *Personality disorders and the five-factor model of personality* (2nd ed., pp. 325-350). Washington, DC: American Psychological Association.

Lynam, D. R., Caspi, A., Moffitt, T. E., Raine, A., Loeber, R., & Stouthamer-Loeber, M. (2005). Adolescent psychopathy and the big five: Results from two samples. *Journal of Abnormal Child Psychology*, *33*, 431-443.

Lynam, D. R., & Derefinko, K. (2005). Psychopathy and personality. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 133-155). New York: Guilford.

Lynam, D. R., & Gudonis, L. (2005). The development of psychopathy. *Annual Review of Clinical Psychology*, *1*, 381-408.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, *1*, 173-181.

Meehl, P. E. (1975). Genes and the unchangeable core. *Voices: The Art and Science of Psychotherapy*, *10*, 25-35.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*, 40-48.

Morrow-Howell, N. (1994). The M word: Multicollinearity in multiple regression. *Social Work Research*, *18*, 247-251.

Newman, J. P., & Schmitt, W. A. (1998). Passive avoidance in psychopathic offenders: A replication and extension. *Journal of Abnormal Psychology*, *107*, 527-532.

Newman, J. P., Schmitt, W. A., & Voss, W. (1997). The impact of motivationally neutral cues on psychopathic individuals: Assessing the generality of the response modulation hypothesis. *Journal of Abnormal Psychology*, *106*, 563-575.

O'Brien, B. S., & Frick, P. J. (1996). Reward dominance: Associations with anxiety, conduct problems, and psychopathy in children. *Journal of Abnormal Psychology*, *24*, 223-240.

Pardini, D., Lochman, J. E., & Frick, P. J. (2003). Callous/unemotional traits and social-cognitive processes in adjudicated youths. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 364-371.

Patrick, C. J. (1995, fall). Emotion and temperament in psychopathy. *Clinical Science*, pp. 5-8.

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research*, *39*, 303-328.

Poulin, F., & Boivin, M. (2000). Reactive and proactive aggression: Evidence of a two-factor model. *Psychological Assessment*, *12*, 115-122.

Price, J. M., & Dodge, K. A. (1989). Reactive and proactive aggression in childhood: Relations to peer status and social context dimensions. *Journal of Abnormal Child Psychology*, *17*, 455-471.

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., et al. (2006). The Reactive-Proactive Aggression (RPQ) Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, *32*, 159-171.

Smithmyer, C. M., Hubbard, J. A., & Simons, R. F. (2000). Proactive and reactive aggression in delinquent adolescents: Relations to aggression outcome expectancies. *Journal of Clinical Child Psychology*, *29*, 86-93.

Verona, E., Patrick, P. J., & Joiner, T. E. (2001). Psychopathy, antisocial personality, and suicide risk. *Journal of Abnormal Psychology*, *110*, 462-470.

Donald R. Lynam is professor of psychology at Purdue University. His primary research interests include developmental models of antisocial behavior, the role of individual differences in deviance, the early identification of chronic offenders, and psychopathy at both the juvenile and adult levels.

Rick H. Hoyle is research professor of psychology and neuroscience at Duke University, where he serves as associate director of the Center for Child and Family Policy; director of the Office of Data, Methods, and Research Facilities in the Social Science Research Institute; and director of the Data Core in the Transdisciplinary Prevention Research Center. His substantive research focuses on personality and social influences on self-control. His methodological research focuses on latent variable models of statistical mediation and moderation. He is editor of the *Journal of Social Issues* and associate editor of *Self and Identity*.

Joseph P. Newman is professor and chair of psychology at the University of Wisconsin–Madison. His research examines the psychological processes that contribute to the dysregulation of behavior, emotion, and cognition. Most of this research has focused on syndromes of disinhibition such as psychopathy (antisocial personality disorder), conduct disorder, aggression, and impulsivity. However, his research also examines how emotionality short-circuits cognitive processing and engenders dysregulation in people with high anxiety.