



## Further evidence of the Narcissistic Personality Inventory's validity problems: A meta-analytic investigation—Response to Miller, Maples, and Campbell (this issue)

Seth A. Rosenthal<sup>a,\*</sup>, R. Matthew Montoya<sup>b</sup>, Leigh E. Ridings<sup>b</sup>, Stacey M. Rieck<sup>b</sup>, Jill M. Hooley<sup>c</sup>

<sup>a</sup> Center for Public Leadership, Harvard Kennedy School, Harvard University, United States

<sup>b</sup> Department of Psychology, University of Dayton, United States

<sup>c</sup> Department of Psychology, Harvard University, United States

### ARTICLE INFO

#### Article history:

Available online 6 August 2011

#### Keywords:

Narcissism  
Narcissistic Personality Inventory (NPI)  
Psychological health  
Self-esteem  
Aggression  
Grandiose narcissism  
Pathological narcissism  
Narcissistic Personality Disorder (NPD)  
Normal narcissism  
Meta-analysis

### ABSTRACT

Miller, Maples, and Campbell (this issue) present evidence that Rosenthal and Hooley's (2010) concerns regarding the Narcissistic Personality Inventory's (NPI) relation to psychological health may be unwarranted. To resolve this issue empirically, we conducted a meta-analysis ( $k = 54$ ,  $N = 38,932$ ). Meta-analytic results revealed that a subset of NPI items were indeed problematic; items that function poorly at differentiating narcissists from non-narcissists accounted entirely for the NPI's connection to psychological health. These items were also strongly associated with self-esteem, but unrelated to aggression/anger. In contrast, the remaining NPI items were unrelated to psychological health, but associated with aggression/anger. We conclude that although the NPI measures narcissism, its poorest functioning items also link it to outcomes unrelated to narcissism.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

Miller, Maples, and Campbell (this issue) have offered a number of important empirical and theoretical challenges to the critiques presented by Rosenthal and Hooley (2010) regarding the form and use of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979, 1981). Miller et al. dispute Rosenthal and Hooley's contention that the NPI, as a *measure*, shares an erroneous relation with psychological health (e.g., happiness, optimism) that is not indicative of the true relation between the *construct* of narcissism<sup>1</sup> and psychological health. Rosenthal and Hooley posited that this problem is caused by the poor discriminant validity of some NPI items; these items simultaneously measure narcissism and other constructs. Rosenthal and Hooley provided evidence that only the NPI items that did the poorest job of differentiating narcissists from non-narcissists accounted for the NPI's positive relation to psychological health. The remaining NPI items—those that did a good job

of differentiating narcissists from non-narcissists—were unrelated to psychological health. Rosenthal and Hooley concluded that the relation between the NPI and psychological health was the result of variance in the NPI that was more closely related to self-esteem than to narcissism. In the current article, we concentrate on the empirical findings reported by Miller et al. that are most relevant to challenging Rosenthal and Hooley's results and address the main theoretical critiques put forth by Miller et al.

From a theoretical perspective, Miller and colleagues suggest that critiques of the NPI such as that put forth by Rosenthal and Hooley (2010) “can be boiled down to one basic question: Is the NPI a valid measure of narcissism?” (Miller & Campbell, 2011, p. 146). Miller and colleagues present evidence that the answer is “yes.” They cite numerous NPI-based studies that indicate that the scale is related to a range of important and theoretically intuitive narcissistic outcomes. Examples include reduced romantic commitment (Campbell & Foster, 2002), pathological gambling (Lakey, Rose, Campbell, & Goodie, 2008), the greedy exploitation of natural resources (Campbell, Bush, Brunell, & Shelton, 2005), and numerous others (see Miller & Campbell, 2010; Miller, Widiger, & Campbell, 2010). Miller and colleagues also note that the NPI converges well with other measures of narcissism (Miller, Gaughan, Pryor, Kamen, & Campbell, 2009; see also Maxwell, Donnellan, Hopwood, & Ackerman, 2011). If one were to agree with Miller and colleagues that the validity of the NPI has already been properly established, contrasting the NPI (as a measure) to

\* Corresponding author.

E-mail address: [seth\\_rosenthal@hks.harvard.edu](mailto:seth_rosenthal@hks.harvard.edu) (S.A. Rosenthal).

<sup>1</sup> Throughout this article, the term “narcissism” refers specifically to *grandiose narcissism* (a tendency toward behaviors that are overtly grandiose, entitled, exploitative, etc.), unless otherwise specified. This is distinct from *vulnerable narcissism* (a feeling of specialness combined with a tendency toward shame, emptiness, and social avoidance; see Cain, Pincus, & Ansell, 2008; Dickenson & Pincus, 2003; Pincus et al., 2009). Thus, our hypothesis is that the NPI is problematic as a *measure of grandiose narcissism* in its prediction of psychological health.

narcissism (as a construct), as proposed by Rosenthal and Hooley, would be a needlessly redundant exercise.

In contrast, others argue that the answer to the question of the NPI's validity is "no" (e.g., Cain et al., 2008; Pincus & Lukowitsky, 2010). They posit that the NPI measures relatively normative, healthy characteristics that are discontinuous from the construct referred to as "narcissism" in clinical theory. This discontinuity is indicated by the NPI's robust positive relation with self-esteem and indicators of psychological health such as happiness (Rose, 2002), optimism and life satisfaction (Brown, Budzek, & Tamborski, 2009), and personal and couple well-being (Sedikides, Rudich, Gregg, Kumashiro, & Rusbult, 2004). Further evidence is provided by the NPI's *negative* relation to multiple forms of psychological distress such as pessimism, depression, sadness, anxiety, and neuroticism (Brown et al., 2009; Sedikides et al., 2004). Because of "criterion" problems such as these, some critics suggest that attempts to integrate NPI-based research into the broader narcissism literature may actually "hamper development of a cohesive knowledge base" about narcissism (Pincus & Lukowitsky, 2010, p. 422).

In our view, both propositions are correct to some degree; the most accurate answer to the question of the NPI's validity is "yes and no" (see also Ackerman et al., 2011, for a similar viewpoint). This paradoxical answer results from a split, roughly speaking, within the NPI itself. On the one hand, the scale contains numerous items that measure narcissism well. These items positively impact the NPI's validity. On the other hand, as noted earlier, the NPI also contains items with poor discriminant validity. These items do not differentiate adequately (a) between narcissists and non-narcissists or (b) between narcissism and normative near-neighbor constructs such as explicit self-esteem. Notably, some of these items also appear to have content that is related only obliquely, if at all, to core narcissistic characteristics such as "grandiosity, entitlement, and exploitativeness" (see Maples et al., 2010, p. 560). Rosenthal and Hooley (2010) presented data indicating that poorly discriminating items accounted entirely for the NPI's relation to indicators of psychological health such as happiness and optimism. The same items were also *unrelated* to the quintessentially narcissistic characteristics of aggression and anger (e.g., Bushman & Baumeister, 1998; Papps & O'Carroll, 1998; Reidy, Foster, & Zeichner, 2010; Rhodewalt & Morf, 1998; Twenge & Campbell, 2003). Rosenthal and Hooley referred to such relatively poor items as *NPI-X* items. The clearest examples of *NPI-X* items included "I am assertive" and "I would prefer to be a leader."

Rosenthal and Hooley (2010) used three methods to identify *NPI-X* items: (a) Item response theory (IRT) analysis conducted by O'Shea and Gustafson (1999), (b) expert clinical researchers' validity ratings of each item's text, and (c) traditional subscale divisions developed through exploratory factor analyses (EFA; Emmons, 1987). In each case, only the poorer discriminating items (i.e., *NPI-X*) accounted for the NPI's counterintuitive positive relation with psychological health. With the *NPI-X* items removed, the NPI's remaining items were, as hypothesized, generally unrelated to psychological health and positively related to aggression/anger.

Rosenthal and Hooley (2010) also indicated that the NPI contains numerous items with satisfactory discriminant validity; these items function well at differentiating narcissists from non-narcissists. They predict typically narcissistic outcomes such as those enumerated earlier. The clearest examples of well-functioning *NPI* items (which were referred to as *NPI-N* items) included "I insist on getting the respect that is due me" and "I will never be satisfied until I get all that I deserve." Rosenthal and Hooley found that these items were related to aggression/anger and to a benchmark measure of narcissism. They concluded that these items link the NPI as a whole to "the kinds of negative outcomes more traditionally associated with narcissism" (p. 454). It is noteworthy that the

NPI can also be a valid predictor of some *positive* outcomes theoretically associated with narcissism, such as improved performance under pressure for the sake of self-enhancement (Wallace & Baumeister, 2002).

### 1.1. Data considerations

In response to Rosenthal and Hooley (2010), Miller et al. (this issue) reported two datasets that directly addressed the relation of the *NPI-X* and *NPI-N* subscales to psychological health (see, in particular, their Table 2 and the Neuroticism and Positive Emotions data in their Table 3). Most of Miller et al.'s results offer evidence that contradicts Rosenthal and Hooley's main findings. For instance, Miller et al. found that the *NPI-X* and *NPI-N* subscales largely shared an equally strong positive relation with self-esteem, whereas Rosenthal and Hooley found that the *NPI-N* items were only weakly related to self-esteem. Miller et al. also found that removing the *NPI-X* items did not significantly reduce the composite NPI's positive relation to psychological health. In contrast, Rosenthal and Hooley found that this step essentially eliminated that relation. In general, Miller et al. concluded that Rosenthal and Hooley's divisions of the NPI merely highlighted fairly trivial differences "of degree not kind" between various NPI items and subscales (p. ##).

In assessing the divergence of the findings of Rosenthal and Hooley (2010) and Miller et al. (this issue), we find it difficult, if not impossible, to attribute the distinctions to anything but sampling variation. This is because the disparities occurred in relations between identical (or nearly identical) self-report inventories that participants completed in non-experimental contexts. In other words, there were no issues of study design, experimenter bias, theoretical interpretation, etc., that were likely to account for the contradictory findings. Because the differences appear to be a purely empirical issue, we believe that the best way to understand which datasets (if any) are most representative of the population is with a meta-analytic investigation of relevant studies. Accordingly, we conducted a meta-analysis ( $k = 54$ ;  $N = 38,932$ ) to resolve the debate over the relation of the three sets of *NPI-X* and *NPI-N* subscales to psychological health as well as self-esteem, benchmarks of grandiose narcissism, and aggression/anger.

## 2. Method

### 2.1. Data collection

We first searched electronic databases (e.g., PsycINFO) to identify articles published using the NPI since 2002. We then sent e-mail requests for raw data to the corresponding authors of those articles, with follow-up emails when necessary. Second, we cross-checked our list of authors with two recent NPI-based meta-analyses (Okada, 2009; Twenge, Konrath, Foster, Campbell, & Bushman, 2008), and sent e-mail requests to the corresponding authors of articles not already identified through the database search. Third, we sent requests for data to multiple listservs (e.g., Society for Personality and Social Psychology, Society for Research in Psychopathology). Finally, we identified published narcissism research articles for which we had not received raw NPI data, but which nonetheless reported correlations between our secondary (i.e., non-NPI) variables. Data collection ended on February 15th, 2011.

The data collection process resulted in 54 datasets and information on 38,932 individual participants. Sample sizes ranged from 49 to 18,989 ( $M = 720.96$ ,  $Median = 190.50$ ,  $SD = 2683.37$ ). Of the 54 datasets, 33 had been published, 3 were unpublished dissertations or Master's theses, and 18 were otherwise unpublished datasets. To be included in the *NPI-X*/*NPI-N* portion of the analyses, we required that a database contain, at minimum, raw item-by-item

**Table 1**

Correlations between composite NPI, criterion variables, and NPI subscales.

Criterion	k	N	Composite NPI		NPI subscale									
			Benchmark		Self-esteem	Grandiose narcissism	IRT-X	IRT-N	Expert-X	Expert-N	EFA-X	EFA-N		
Self-esteem	43	36,092	.30*	(.32, .27)	–	–.03 (.01, –.07)	<b>.43*</b> (.45, .41)	.13* (.15, .11)	<b>.31*</b> (.33, .29)	.25* (.27, .23)	<b>.31*</b> (.33, .29)	.24* (.26, .22)		
Grandiose narcissism	7	1,950	.37*	(.43, .30)	–.03 (.01, –.07)	–	.25* (.30, .20)	<b>.44*</b> (.49, .39)	.21* (.26, .16)	<b>.45*</b> (.50, .40)	.21* (.26, .16)	<b>.43*</b> (.48, .38)		
Psychological health	17	9,704	.15*	(.18, .11)	.63*	(.66, .59)	–.20*	(–.26, –.13)	<b>.21*</b> (.24, .18)	.02 (.05, –.01)	<b>.13*</b> (.16, .10)	.08* (.11, .05)	<b>.15*</b> (.18, .12)	.09* (.12, .06)
Aggression/Anger	8	20,332	.16*	(.20, .11)	–.29*	(–.24, –.33)	.24*	(.29, .19)	.03 (.08, –.02)	<b>.29*</b> (.34, .24)	.11* (.16, .06)	<b>.21*</b> (.26, .16)	.07* (.12, .02)	<b>.21*</b> (.26, .16)

\*  $p < .05$ .  $k$  and  $N$  refer to the number of studies and sample sizes for the correlations between the criterion variables and the Composite NPI, as well as the NPI-X and NPI-N subscales. 95% confidence intervals are reported in parentheses. Bolded NPI subscale correlations are of a significantly greater magnitude than are the correlations for their respective paired subscales,  $p < .05$ . The correlations between the paired NPI-X and NPI-N subscales are IRT,  $z = .67$ ; Expert,  $z = .73$ ; EFA,  $z = -.61$ .

NPI data (from the 37- or 40-item version of the NPI) and a validated measure of self-esteem. The exception to the raw data requirement was Miller et al. (this issue), for whom we used the results reported in their article.

## 2.2. Data coding

### 2.2.1. Narcissistic Personality Inventory

For raw data from the 37-item version of the NPI, the NPI-X and NPI-N subscales were computed as outlined in Rosenthal and Hooley (2010, Table 1): Thirty-seven items were used to code the Expert- and EFA-based subscales, and 31 items were used to code the IRT-based subscales. For raw data from the 40-item NPI, all subscales were computed using 31 items (see Miller et al., this issue). It is important to note that identical sets of 31 items were used to code the IRT-based subscales for both the 37- and the 40-item versions of the NPI.

### 2.2.2. Self-esteem

All datasets but one measured self-esteem using the Rosenberg Self-Esteem Scale (Rosenberg, 1965). The remaining dataset used the Single-Item Self-Esteem Scale (Robins, Hendin, & Trzesniewski, 2001).

### 2.2.3. Psychological health

We included as indicators of psychological health measures that assessed a positive psychological orientation (e.g., happiness, life satisfaction) or a negative psychological orientation (e.g., depression). Measures of a positive psychological orientation included the Oxford Happiness Questionnaire (Hills & Argyle, 2002), the Satisfaction With Life Scale (Diener, Emmons, Larson, & Griffin, 1985), Yarkoni's (2010) abbreviated version of the Well-Being subscale of the Multidimensional Personality Questionnaire (Tellegen & Waller, 1994), and the Presence subscale of the Meaning in Life Questionnaire (Steger, Frazier, Oishi, & Kaler, 2006). Measures of a negative psychological orientation, which were reverse-scored, included the Global Severity Index of the Brief Symptom Inventory (Derogatis & Melisaratos, 1983) and a number of depression measures including the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the Center for Epidemiologic Studies Depression Scale (Radloff, 1977), the Hamilton Rating Scale for Depression (Hamilton, 1960), and the Depressiveness facet score of the NEO Personality Inventory–Revised (Costa & McCrae, 1992).

### 2.2.4. Grandiose narcissism

We included validated measures of grandiose narcissism to serve as a benchmark. Measures included aggregated scores of the Pathological Narcissism Inventory's (Pincus et al., 2009) three grandiose subscales (i.e., Exploitativeness, Self-Sacrificing Self-Enhancement, and Grandiose Fantasy; see Wright, Lukowitsky, Pincus, & Conroy, 2010), the Childhood Narcissism Scale (Thomaes, Stegge, Bushman, Olthof, & Denissen, 2008), the narcissism

composite of the Schedule for Nonadaptive and Adaptive Personality (Clark, 1993), and a number of measures directly related to the criteria for Narcissistic Personality Disorder (NPD) enumerated in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R, APA, 1987; DSM-IV, APA, 1994). These included the NPD section of the Personality Diagnostic Questionnaire-4 (Hyler, 1994), the NPD sections of the Structured Clinical Interview for DSM-IV Personality Disorders and the Personality Questionnaire for the SCID-II (First, Gibbon, Spitzer, Williams, & Benjamin, 1997), and a battery of clinician-administered DSM-based assessments (see Miller, Campbell, & Pilkonis, 2007).

### 2.2.5. Aggression/anger

An aggressive and angry orientation toward others was operationalized using the Aggression Questionnaire (AQ; Buss & Perry, 1992), the Angry Hostility facet score of the NEO Personality Inventory–Revised (Costa & McCrae, 1992), the Proactive Aggression subscale of the Reactive–Proactive Aggression Questionnaire (Raine et al., 2006), and a task in which participants assigned a potentially punitive grade to fabricated written comments about an essay (see Bushman et al., 2009).<sup>2</sup>

### 2.2.6. Other variables

For each dataset, we coded basic descriptive information and additional variables for sensitivity and moderator analyses. These variables included: source type (journal, edited volume, thesis or dissertation, and unpublished dataset), sample size, sample composition (college students, adults, or school children), NPI version (37- or 40-item), language (English, German), country (United States, Germany, Sweden), NPI question type (dichotomous versus continuous), method of administration (on-line, in person), author(s) and full citation, and year of publication.

## 3. Statistical methods

### 3.1. Effect size

The effect size index was Fisher's  $z$  (Fisher, 1928), calculated such that greater values indicate a stronger relation between the measures of interest. If the 95% confidence interval for the effect size includes zero, it indicates that there is no significant relation between two constructs.

### 3.2. Random-effects model

When conducting a meta-analysis, researchers select either a fixed-effects model or random-effects model (Field, 2001; Hunter

<sup>2</sup> Although aggression and anger are separable constructs, the authors of the AQ (Buss & Perry, 1992) considered a composite that combines elements of both to be valid standard by which to measure aggression. Also, past narcissism research has investigated aggression and anger as a meaningfully paired outcome (e.g., Pappas & O'Carroll, 1998; Twenge & Campbell, 2003).

& Schmidt, 2000). We selected a random-effects model for two reasons. First, we were interested in making unconditional inferences that generalized to the hypothetical population of all studies that could exist rather than simply to the studies included in the present sample (Hedges & Vevea, 1998). Conclusions drawn from random-effects models generalize to contexts of all possible operational definitions, whereas conclusions drawn from fixed-effects models generalize only to contexts involving the operational definitions used by the meta-analyzed studies.

The second reason for selecting a random-effects model was the tendency for data to violate the assumption of homogeneity (Field, 2001; National Research Council, 1992). When that assumption is violated, fixed-effects models underestimate the standard errors of parameter estimates and inflate the Type I error rate. For example, Monte Carlo simulations indicated that the Type I error rate (which is usually set at .05) ranged between .43 and .80 in heterogeneous fixed-effects models (Field, 2003). We computed a *Q*-statistic for each analysis to test the assumption of homogeneity. A significant *Q*-statistic indicates heterogeneity by detecting the additional uncertainty not captured by the fixed-effects analyses. In the case of a significant *Q*-statistic, the random-effects model is the appropriate test of meta-analytic hypotheses.

#### 4. Results

Before we estimated the means for any of the variables of interest, we computed a *Q*-statistic of the null hypothesis that it is plausible the true variance component was zero. For each mean estimated and for each analysis computed, the *Q*-test was significant. As a result, we used random-effects models for all of the analyses reported below.

The resulting population effect sizes were interpreted using Cohen's (1988) suggestion that an *r* of at least .10 be labeled a "small effect," an *r* of at least .30 a "medium effect," and an *r* of at least .50 a "large effect."

##### 4.1. Self-Esteem and grandiose narcissism

The first row of Table 1 reports the correlations between self-esteem and the composite NPI, benchmark grandiose narcissism, and the NPI-X and NPI-N subscales. Consistent with past findings, the composite NPI was positively correlated with self-esteem. The magnitude of this correlation ( $z = .30$ ) was equivalent to that generally reported in the NPI literature (see Miller & Campbell, 2011). This provides evidence that on the whole, the current dataset was representative and generalizable.

Inspection of the Expert- and EFA-based divisions of the NPI indicated that the NPI-X and NPI-N subscales were each positively correlated with self-esteem. The correlations of the NPI-X subscales to self-esteem were significantly stronger than were NPI-N correlations with self-esteem. Strongest support for a difference between the NPI-X and NPI-N items as they relate to self-esteem was evident in the IRT-based division of the scale.

In distinction to the NPI subscales' positive relations with self-esteem, if the benchmark measures of grandiose narcissism are to be viewed as exhibiting the "accurate" relation between narcissism and self-esteem, the results revealed that narcissism and self-esteem were unrelated ( $z = -.03$ ). In contrast, the link between the composite NPI and self-esteem was robustly positive ( $z = .30$ ), and was comparable to the relations between self-esteem and the NPI-X items alone ( $z$ s = .31–.43).

The second row of Table 1 presents the correlations between grandiose narcissism and the composite NPI, as well as the NPI-X

and NPI-N subscales. The composite NPI was more strongly related to grandiose narcissism ( $z = .37$ ) than to self-esteem ( $z = .30$ ),  $t = 2.33$ ,  $p < .05$ , although both relations were notably positive. This generally supports the claim by Rosenthal and Hooley (2010) that the composite NPI measures both narcissism and self-esteem, although in contrast to their findings, it does measure narcissism better than it measures self-esteem.

In terms of the NPI subscales, each NPI-X and NPI-N subscale had a significant positive relation with grandiose narcissism. This provides support for the contention of Miller et al. (this issue) that the NPI-X items do measure narcissism. However, each NPI-X subscale had a weaker relation with grandiose narcissism, and a stronger relation with self-esteem than did its paired NPI-N subscale. Further, each of the NPI-X subscales was more strongly related to self-esteem than to grandiose narcissism, whereas the opposite was true for the NPI-N subscales. These results provide support for the case made by Rosenthal and Hooley (2010) that the NPI-X items have poor discriminant validity in that they are more closely related to self-esteem than to narcissism.

##### 4.2. Psychological health and aggression/anger

The third row of Table 1 indicates that the composite NPI was positively related to psychological health ( $z = .15$ ). This significant positive relation accords with past theory (e.g., Campbell, 2001) and research investigating the positive link between NPI and psychological health (e.g., Brown et al., 2009; Okada, 2009; Rose, 2002; Sedikides et al., 2004). The relation is consistent with, although much weaker than, the large positive relation between self-esteem and psychological health ( $z = .63$ ). However, it is inconsistent with the negative relation between grandiose narcissism and psychological health ( $z = -.20$ ). This supports Rosenthal and Hooley's (2010) assertion that the NPI's relation with psychological health is at-odds with the "true" relation between the construct of narcissism and psychological health.

The fourth row of Table 1 indicates that the composite NPI and grandiose narcissism shared similar positive relations with aggression/anger. These relations are consistent with those typically reported in the literature. They lend support to the agreed-on position of Rosenthal and Hooley (2010) and Miller et al. (this issue) that the composite NPI covaries with typically narcissistic constructs such as aggression/anger. In contrast, self-esteem shared a negative correlation with aggression/anger (see Donnellan, Trzesniewski, Robins, Moffitt, & Caspi, 2005 versus Bushman et al., 2009).

In terms of the NPI subscales, each of the NPI-X subscales was more strongly related to psychological health than was its paired NPI-N subscale, whereas each NPI-N subscale was more strongly related to aggression/anger than was its paired NPI-X subscale. For the Expert- and EFA-derived solutions, the NPI-N relations with psychological health, and the NPI-X relations with aggression/anger were also significant.

However, the IRT-based solution provided a different, and more definitive, outcome. As noted above, the IRT-X subscale was more strongly related to psychological health than was the IRT-N subscale. Additionally, the IRT-N subscale was unrelated to psychological health ( $z = .02$ ). Likewise, the IRT-N subscale was related to aggression/anger whereas the IRT-X subscale was unrelated to aggression/anger ( $z = .03$ ). The IRT-based results provide strong support for Rosenthal and Hooley's (2010) main hypothesis, in that particular NPI items, which were independently identified as the scale's poorest internally functioning items, accounted entirely for the NPI's relation with psychological health. Importantly, these same items were unrelated to a



theoretically critical correlate of narcissism (i.e., aggression/anger).<sup>3,4</sup>

#### 4.3. Sensitivity analyses

We conducted additional analyses to determine whether the relations between the variables of interest were affected by assessment tool or demographic characteristics. Results indicated that assessment tool generally did not affect the key relations. For instance, the NPI did not differ in its relation with the measures used to assess grandiose narcissism,  $\chi^2(3) = 0.83, p = .84$ . The measures of grandiose narcissism did differ in their relations with self-esteem,  $\chi^2(5) = 14.83, p < .05$ . However, the corresponding correlations between each grandiose narcissism measure and self-esteem were all near zero or negative, ranging from  $z = .01$  (Grandiose Composite of the Pathological Narcissism Inventory) to  $z = -.21$  (Structured Clinical Interview for DSM-IV Personality Disorders).

Relations were also generally unaffected by demographic characteristics. For example, the relation between the composite NPI and psychological health was not affected by NPI version (37- versus 40-item),  $\chi^2(1) = 0.32, p = .57$ , or the language used for the assessments,  $\chi^2(1) = 0.01, p = .92$ . The same was true for the NPI's relation with measures of aggression/anger, which also were unaffected by NPI version,  $\chi^2(1) = 2.64, p = .10$ , and language,  $\chi^2(1) = 0.76, p = .38$ .

We also investigated whether the key IRT-based subscale findings differed according to the measure used to assess psychological health or aggression/anger. Results indicated that the null relation of the IRT-N subscale to psychological health was unaffected by measure of psychological health,  $\chi^2(5) = 2.80, p = .73$ . An additional analyses tested whether the null relation of the IRT-N subscale to psychological health held separately for measures that were positively oriented (e.g., happiness) and negatively oriented (e.g., depression). Results indicated that the IRT-N subscale was unrelated to both the positively oriented measures ( $z = .03; k = 9$ ) and the (reverse-scored) negatively oriented measures ( $z = .01; k = 8$ ). Further, there was no difference in the magnitude of these correlations,  $\chi^2(1) = 0.53, p = .46$ .

With respect to aggression/anger, all eight of the studies that examined the relation between the IRT-X subscale and aggression/anger used the Aggression Questionnaire. However, because additional aggression/anger measures were included in the estimate of the relation between the benchmark grandiose narcissism composite and aggression/anger, a supplemental analysis was conducted. The correlation between the grandiose narcissism benchmark and aggression/anger was not affected by the measure used to assess aggression/anger,  $\chi^2(3) = 4.47, p = .21$ .

## 5. Discussion

The meta-analytic results provided clear evidence that the answer to Miller and Campbell's (2011) question—"is the NPI a valid

measure of narcissism?" (p. 146)—truly is "yes and no." The reason for this "split decision" was most clearly illustrated by the IRT-based division of the scale. The answer is "yes" because the NPI contains the IRT-N items that had a robust positive relation to other measures of grandiose narcissism, a smaller relation to self-esteem, and were related to aggression/anger but not to psychological health. These findings generally lend support to the emphasis placed by Miller et al. (this issue) on evidence of the NPI's external validity.

However, the results also indicated that the answer to the validity question is "no." The NPI contains the IRT-X items, and these had a nearly opposite set of external relations relative to the IRT-N items: They were positively related to psychological health, but unrelated to aggression/anger. They were also more strongly related to self-esteem ( $z = .43$ ) than to benchmark measures of grandiose narcissism ( $z = .25$ ). Taken together, it appears that whereas individuals who endorse IRT-N items are likely to exhibit specifically narcissistic characteristics, individuals who endorse the IRT-X items will likely exhibit a poorly differentiated mix of narcissistic and non-narcissistic characteristics.

Overall, the meta-analysis affirmed the generalizability of the IRT-based results reported by Rosenthal and Hooley (2010) that were challenged by Miller et al. (this issue). They also supported the key assertions made by Rosenthal and Hooley. They indicated that a small subset of the NPI's items account entirely for the scale's positive relation with psychological health. This result would not necessarily be noteworthy in isolation, particularly if there was no other objective evidence that the IRT-X items differed from the rest of the NPI's items or were problematic. However, the IRT-X items also failed to demonstrate the positive correlation with aggression/anger that was exhibited by the composite NPI, the IRT-N items, and the grandiose narcissism composite measure. And most importantly, the IRT-X items were identified *a priori* using a theoretically-based objective method; they are the items in the NPI that do the poorest job of differentiating narcissists from non-narcissists. Because of this, it is fairly clear that the reason for the paradoxical results is that the IRT-X items measure narcissism *as well as other constructs*. And it is this additional variance that leads the IRT-X subscale, as well as the composite NPI, to correlate positively with psychological health.

#### 5.1. Should poorly discriminating NPI items be retained regardless?

Miller and colleagues (Miller & Campbell, 2011; Miller et al., this issue) have contended that even the NPI's most controversial items (e.g., assertiveness- and leadership-based items) should be viewed as part of a well-validated narcissism measure. And according to our meta-analytic results, even the poorest performing NPI items (i.e., IRT-X) shared a positive relation with benchmark measures of narcissism ( $z = .25$ ). Further, the text of most of the IRT-X items does in fact bear some relation (sometimes even a prototypical relation; see Lynam & Widiger, 2001; Samuel & Widiger, 2004) to narcissism.

For these reasons, Rosenthal and Hooley (2010) may have been imprecise when suggesting that NPI-X items "bear little resemblance to the narcissism construct" (p. 461). Characteristics such as assertiveness (see Lynam & Widiger, 2001; Samuel & Widiger, 2004, 2008) and leadership (see Rosenthal & Pittinsky, 2006) are related to narcissism, although it is less apparent that they represent core aspects of narcissism such as grandiosity, entitlement, exploitativeness, etc.

The fact that the IRT-X items have some relation with narcissism does not address the problem of their poor discriminant validity, however. In other words, one must also ask what *else*, other than narcissism, these items might measure. Are the IRT-X items confounded with other constructs? In light of this question,

<sup>3</sup> There was concern that the key results were affected disproportionately by one particularly large sample ( $N = 18,989$ ). This sample did not include psychological health data. However, a supplemental analysis indicated that removal of this database did not affect the key anger/aggression finding; with this large sample removed, the correlation between the IRT-X subscale and aggression/anger remained non-significant ( $z = .02$ ).

<sup>4</sup> For reference, we have prepared an on-line supplement that examines the NPI's external relations using established factor-based solutions and a shortened version of the NPI. Table 2 includes correlations of the NPI subscales established by Emmons (1987), Raskin and Terry (1988), Corry, Merritt, Mrug, and Pamp (2008), and Ackerman et al. (2011), as well as the 16-item version of the scale (Ames, Rose, & Anderson, 2006) with self-esteem, grandiose narcissism, psychological health, and aggression/anger. Further, in response to data presented by Miller et al. (this issue), we have also included an examination of the NPI's relation to the Five Factor Model of Personality (FFM;  $k = 11; N = 29,587$ ). Table 3 includes correlations of the composite NPI, self-esteem, grandiose narcissism, and the NPI-X and NPI-N subscales with the FFM personality traits. Please see Appendix A. Supplementary material.

the meta-analytic results indicated that the IRT-X items shared more than 10 times the variance with self-esteem as did the IRT-N items. More strikingly, the IRT-X items were more strongly related to self-esteem than to benchmark measures of grandiose narcissism. This provides strong evidence that Rosenthal and Hooley (2010) were correct in asserting that these items do a better job of measuring self-esteem than narcissism.

Most noteworthy is the finding, mentioned earlier, that without the IRT-X items, the NPI did not exhibit a positive relation with psychological health. This relation was entirely dependent on the variance produced by the NPI's poorest discriminating items. This meta-analytic finding is at-odds with what Miller et al. (this issue) found in their datasets. Accordingly, we believe it is crucial to consider all NPI-based findings both with and without the IRT-X items. In situations in which this procedure produces conflicting results, careful consideration in light of the current analyses must be given to favoring the results produced by the IRT-N items (see Ackerman et al., 2011, for an extended discussion of paradoxical NPI-based results).

### 5.2. Is it still appropriate to "suggest caution" when using the NPI?

The findings reported by Rosenthal and Hooley (2010) should not be interpreted as casting "doubt on the validity of the majority [of NPI-based] empirical work on narcissism" (see Miller et al., this issue, p. ##; emphasis added). It is clear that many valid inferences about narcissism can be and have been made using the NPI.

Nonetheless, increased caution when using the NPI should be applied in two particular circumstances. First, as we have discussed at length, we would advise caution when results based on the NPI's poorer performing items are inconsistent with results based on its better performing items. In such cases, all findings should be reported, interpreted, and resolved carefully. Reference should be made to which of the paradoxical findings (a) is produced by the NPI's better discriminating items, (b) relates more closely to core aspects of narcissism (e.g., grandiosity, entitlement, exploitativeness, etc.), and (c) is replicable using other validated narcissism measures.

Second, although not addressed directly by the current data, we continue to advocate caution in making inferences about narcissism if they are fully dependent on the NPI's overlap with self-esteem (e.g., narcissism's positive relation to psychological health; see Brown et al., 2009; Rose, 2002; Rosenthal & Hooley, 2010; Sedikides et al., 2004). In contrast to Miller et al. (this issue), we believe that *not* doing so is a cause for concern (e.g., Paulhus, 1998, 2001; Paulhus, Robins, Trzesniewski, & Tracy, 2004; Tracy, Cheng, Robins, & Trzesniewski, 2009). However, we would specify that there are two narrow criteria necessary for triggering this enhanced caution. The first criterion is that the mediation of the NPI-to-outcome link must be full; the NPI's relation to the outcome variable must be nullified, or the direction of its relation reversed, by controlling for self-esteem. In cases where the NPI maintains predictive power above-and-beyond mediation by self-esteem, we would agree that the outcome variable should be described as at least partially predicted by narcissistic variance provided by the NPI (in other words, it is predicted by the construct of narcissism).

The second criterion is that the caution applies only if the mediating variable is not a defining, core characteristic of narcissism. Self-esteem, although linked to the NPI, is not a defining characteristic of narcissism in the manner in which grandiosity, entitlement, exploitativeness, etc. are defining characteristics. Unlike those core narcissistic characteristics, self-esteem is not also *defined* by an intrinsic, reciprocal relation to narcissism (see Johnson, 2004). In other words, one cannot deduce simply from a high self-esteem score whether an individual is or is not likely to be narcissistic.

In contrast, high levels of grandiosity, entitlement, exploitativeness, etc. are intimately involved in determining whether an individual is narcissistic. Because of this distinction, one must interpret mediation by these constructs differently. Full mediation by a self-esteem scale indicates that a zero-order link between the NPI and an outcome is solely the result of variance provided by normative self-esteem. With self-esteem's variance removed, the NPI provides no additional, *specifically narcissistic* variance that predicts the outcome.<sup>5</sup> This was the interpretation provided by Rosenthal and Hooley (2010) of the NPI-based research that suggests that psychological health is predicted by narcissism via its relation to self-esteem (e.g., Rose, 2002; Sedikides et al., 2004). We believe that it is at least as theoretically sound to interpret such data to indicate that psychological health is fully predicted by normative self-esteem, with no independent, specifically narcissistic, variance contributed by NPI scores.

It is crucial for all researchers to have the means to examine the potential for confounds within any scale. We believe this is one of the reasons that Lynam, Hoyle, and Newman (2006) provided greater latitude to interpret partial correlations to "make statements about...a scale" [in contrast to a broader construct] (p. 329). Accordingly, self-esteem can and should be consistently examined and reported as a potentially confounding second construct within the NPI. Any NPI-based relation that is fully accounted for by a measure of normative self-esteem should be interpreted cautiously. In such cases, the burden should be placed on the researcher to provide evidence that the connection of the NPI to the outcome via normative self-esteem also involves characteristics that are specifically attributable to narcissism.

### 5.3. Interpretive issues

#### 5.3.1. IRT versus Expert and EFA

In contrast to the IRT-N subscale, the Expert-N and EFA-N subscales were significantly related to psychological health. This contradicts the results reported by Rosenthal and Hooley (2010), and appears to lend partial support to Miller et al.'s critique that differences between the NPI-N and NPI-X subscales may merely be a matter of degree.

However, it is clear that in each of the three divisions of the scale, the NPI-N subscales accounted for a trivial amount of the variance in psychological health (see Cohen, 1988). Further, the IRT-based division of the NPI may have offered some advantages over the other two methods. By presenting *b* parameter estimates for each NPI item, the IRT analysis indicated how effectively each item differentiated between individuals high and low in narcissism. This provided an objective assessment of the function of each NPI item, whereas the expert ratings provided a subjective assessment of each item. Focusing on each item's relation to the full NPI was also particularly appropriate, relative to examining the scale's internal factor structure. This is because the NPI was initially developed using a unidimensional internal consistency strategy (see Raskin & Hall, 1979; Raskin & Terry, 1988), and it is most often used by researchers in its composite form, ignoring its factor structure (Miller & Campbell, 2008).

The IRT-based results should, however, be regarded with some degree of caution for at least two reasons. First, the analysis that

<sup>5</sup> This contrasts with a situation in which a full mediator is a core characteristic of narcissism. For instance, the NPI may predict certain outcomes that are fully mediated by a measure of psychological entitlement (see Campbell, Bonacci, Shelton, Exline, & Bushman, 2004). In such a case, one would conclude that the outcomes were (a) predicted by the general construct of narcissism as measured by the NPI and (b) were specifically the result of narcissists' sense of entitlement. Because entitlement is a core characteristic of narcissism, the NPI's effect would not be *confounded* by the full mediation of entitlement, but would rather be clarified by it. Thus, no additional caution would be needed when interpreting these results.

Rosenthal and Hooley (2010) used as the framework for constructing the IRT-based subscales (i.e., O'Shea & Gustafson, 1999) has not been peer-reviewed or published. The second is that the *b* parameter dividing line between the IRT-N and IRT-X items was selected subjectively by Rosenthal and Hooley. Because of this, the best functioning IRT-X item may perform almost as well as the worst functioning IRT-N item. It is important to reiterate, however, that the unpublished IRT did not, itself, provide data pertaining to the NPI's relation to psychological health. Instead, this analysis was applied, *a priori*, as a theoretically meaningful framework for dividing the scale. Only after the IRT-based framework was applied to splitting the NPI were the independently collected raw data analyzed to determine the outcomes for the IRT-N and IRT-X subscales.

Ultimately, we would argue that cautions regarding the IRT framework, and/or the failure of the Expert- and EFA-based solutions to produce the precise predicted results, do not diminish the importance of the IRT-based findings. Any theoretically meaningful, empirically determined configuration of items that reveals consequential bias within the NPI signals a serious problem with the scale. The fact that the scale's flaws are most evident using one particular arrangement of its items does not mitigate the concern. The IRT-based results indicate that the NPI items with the poorest internal performance also account fully for its counterintuitive external performance (i.e., its positive relation to psychological health). The IRT-based results suggest that it may be important to place objective consideration of how well NPI items differentiate narcissists from non-narcissists ahead of subjective concerns about items' form (e.g., Should the NPI include items about leadership or assertiveness?; see Miller & Campbell, 2011; Miller et al., this issue) or consideration of the ways in which items converge into particular factors.

### 5.3.2. Benchmark measures of grandiose narcissism

There is considerable debate about whether the measures included in our composite grandiose narcissism measure offer a proper benchmark (a) by which to operationalize grandiose (in contrast to vulnerable) narcissism and/or (b) by which to judge the validity of the NPI. On the one hand, the DSM criteria on which most of these measures are based "entirely or primarily" reflect aspects of grandiose narcissism (Miller & Campbell, 2008, p. 456; see also Miller et al., 2010). On the other hand, Miller and colleagues (Miller, Campbell, Pilkonis, & Morse, 2008; Miller et al., in press) also argue that the actual items, scales, and interviews used to measure the DSM NPD criteria draw on variance related to a *mix* of grandiose and vulnerable narcissism.

Others have provided a more inclusive conceptualization of grandiose narcissism, and thus a more circumscribed definition of vulnerable narcissism. For instance, they suggest that "all [DSM-based] NPD measures assess...narcissistic grandiosity" (Pincus & Lukowitsky, 2010, p. 432; see also Pincus et al., 2009). In contrast, vulnerable narcissism is defined by characteristics that are subjectively different from those enumerated by the NPD criteria, such as hypersensitive emotional responses to ridicule (Hendin & Cheek, 1997) and feelings of shame and worthlessness (Pincus et al., 2009). Empirically, in contrast to the assertion of Miller et al. (this issue) that "high self-esteem...characterizes grandiose narcissism" (p. ##; see also Bosson et al., 2008), Pincus et al. presented evidence that it is possible to measure aspects of grandiose narcissism that do not share a positive correlation with self-esteem (see also Maxwell et al., 2011). Ultimately, we agree with the basic premise that the parameters set by Miller and colleagues for defining grandiose narcissism and differentiating it from vulnerable narcissism are too narrow and too dependent on the NPI and its nomological network as a benchmark.

We note further, that even if one sets aside the debate about grandiose versus vulnerable narcissism, DSM-based and similar

measures of narcissism should still be considered appropriate benchmarks for judging the NPI's validity. The NPI was developed to measure narcissism "as defined by" the DSM criteria (Raskin & Hall, 1979, p. 590). Because of this, the failure of some NPI subscales and items to converge more strongly with DSM-based narcissism measures than with measures of self-esteem likely reflects flaws in the NPI rather than in measures of the DSM NPD criteria on which the NPI is purportedly based (for additional discussion, see Rosenthal & Hooley, 2010).

## 6. Conclusion

As noted by Rosenthal and Hooley (2010), a valid scale should "measure one thing (i.e., the target construct)—and only that thing—as precisely as possible" (Clark & Watson, 1995, p. 315). In light of this recommendation, the current research focused on problems with the NPI's internal validity (i.e., its precision) and its discriminant validity (i.e., its only-that-thing-ness). The meta-analysis indicated that because of validity problems in these two realms, the NPI does not perform optimally, and can sometimes produce misleading results. The NPI clearly does measure narcissism. Miller and colleagues have argued this case convincingly (e.g., Miller & Campbell, 2010; Miller et al., this issue), and the meta-analytic results support this assertion. However, our results also indicate that the NPI does not measure narcissism *exclusively*.<sup>6</sup>

Questions regarding the NPI's validity are usually considered to be theoretical and subjective. Without a single, authoritative definition of narcissism, inconclusive debates over the NPI's validity would likely continue *ad nauseum*. However, the addition of an IRT-based framework can help to triangulate empirically and objectively on a solution. The current meta-analysis demonstrated that the NPI's items that function most poorly at determining who is and is not narcissistic are the *same* items that account for most of the normative, healthy variance measured by the NPI. Because of this, they produce questionable findings about the putative healthy side of narcissism. They also attenuate the strength of the relation between the NPI and typically narcissistic correlates such as aggression/anger. Both outcomes are counterproductive to furthering our understanding of narcissism. Accordingly, we believe that a different basic question—do these specific items contribute to, or detract from, the NPI's validity?—needs to be addressed systematically (see Ackerman et al., 2011; Rosenthal & Hooley, 2010). The current data appear to indicate that on the whole, these items create more problems than they solve, particularly as far as discriminant validity is concerned.

Because of the problematic functioning of some NPI items, we believe that it is imperative for researchers to consider performing NPI-based analyses without these items. In other words, we recommend fine-tuning the NPI just enough so that it no longer generates meaningful variance from individuals who may be assertive and confident, but are also non-grandiose, non-entitled, and non-exploitative, etc. This leaner, and literally meaner, NPI would produce results that are more consistent with the other validated measures of narcissism. It would also, in contrast to the concerns voiced by Miller et al. (this issue), likely *strengthen* most of the results obtained in NPI-based research, only diminishing those associated with the non-narcissism-specific variance that confounds the scale. Finally, it might also help satisfy some of the NPI's critics (e.g., Cain et al., 2008; Pincus & Lukowitsky, 2010) by

<sup>6</sup> This is not terribly surprising, as we find no evidence that the scale's ability to discriminate between narcissism and other near-neighbor constructs was referenced or tested to aid in deciding which of the NPI's original 223 items (see Raskin & Hall, 1979) to retain in its authors' subsequent 80-, 54-, or 40-item versions (see Raskin & Hall, 1981; Raskin & Terry, 1988).



making it easier to integrate the NPI into the broader literature of narcissism theory and research (e.g., Miller & Campbell, 2010).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the on-line version at doi:10.1016/j.jrp.2011.06.004.

## References

References marked with an asterisk indicate studies included in the meta-analysis.

- \*Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What does the Narcissistic Personality Inventory really measure? *Assessment*, 18, 67–87.
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., Revised). Washington, DC: American Psychiatric Association.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40, 440–450.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.
- \*Bosson, J. K. (2001). When conscious and non-conscious self-views clash: Interpersonal and physiological correlates of discrepant implicit and explicit self-esteem. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 61(11-B), 6184–6294.
- \*Bosson, J. K. (2004). [NPI-40, Rosenberg self-esteem, psychological health]. Unpublished raw data.
- \*Bosson, J. K. (2005). Overvalued and ashamed: Considering the roles of self-esteem and self-conscious emotions in covert narcissism. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (2nd ed., New York: Guilford).
- Bosson, J. K., Lakey, C. E., Campbell, W. K., Zeigler-Hill, V., Jordan, C. H., & Kernis, M. H. (2008). Untangling the links between narcissism and self-esteem: A theoretical and empirical review. *Social and Personality Psychology Compass*, 2(3), 1415–1439.
- \*Brown, R. P., Budzek, K., & Tamborski, M. (2009). On the meaning and measure of narcissism. *Personality and Social Psychology Bulletin*, 35, 951–964.
- \*Bushman, B. (2011). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, 75, 219–229.
- \*Bushman, B., Baumeister, R. F., Thomaes, S., Ryu, E., Begeer, S., & West, S. G. (2009). Looking again and harder, for a link between low self-esteem and aggression. *Journal of Personality*, 77, 427–446.
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63, 452–459.
- Cain, N. M., Pincus, A. L., & Ansell, E. B. (2008). Narcissism at the crossroads: Phenotypic description of pathological narcissism across clinical theory, social/personality psychology, and psychiatric diagnosis. *Clinical Psychology Review*, 28, 638–656.
- Campbell, W. K. (2001). Is narcissism really so bad? *Psychological Inquiry*, 12, 214–216.
- Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment*, 83, 29–45.
- Campbell, W. K., Bush, C. P., Brunell, A. B., & Shelton, J. (2005). Understanding the social costs of narcissism: The case of the tragedy of the commons. *Personality and Social Psychology Bulletin*, 31, 1358–1368.
- Campbell, W. K., & Foster, C. A. (2002). Narcissism and commitment in romantic relationships: An investment model analysis. *Personality and Social Psychology Bulletin*, 28, 484–495.
- \*Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, 31, 334–347.
- Clark, L. A. (1993). *Schedule for Nonadaptive and Adaptive Personality (SNAP). Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Corry, N., Merritt, R. D., Mrug, S., & Pamp, B. (2008). The factor structure of the Narcissistic Personality Inventory. *Journal of Personality Assessment*, 90, 593–600.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.
- \*Crocker, J., & Canevello, A. (2008). Creating and undermining social support in communal relationships: The role of compassionate and self-image goals. *Journal of Personality and Social Psychology*, 95, 555–575.
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, 13, 595–605.
- Dickenson, K. A., & Pincus, A. L. (2003). Interpersonal analysis of grandiose and vulnerable narcissism. *Journal of Personality Disorders*, 17, 188–207.
- Diener, E., Emmons, R. A., Larson, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49, 71–75.
- Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., Moffitt, T. E., & Caspi, A. (2005). Low self-esteem is related to aggression, antisocial behavior, and delinquency. *Psychological Science*, 16, 328–335.
- Emmons, R. A. (1987). Narcissism: Theory and measurement. *Journal of Personality and Social Psychology*, 52, 11–17.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161–180.
- Field, A. P. (2003). The problems using fixed effects models of meta-analysis on real-world data. *Understanding Statistics*, 2, 77–96.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *User's guide for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, DC: American Psychiatric Press.
- Fisher, R. A. (1928). *Statistical methods for research workers*. Oxford, England: Stechert.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56–62.
- Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's Narcism Scale. *Journal of Research in Personality*, 31, 588–599.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504.
- Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: A compact scale for the measurement of psychological well-being. *Personality and Individual Differences*, 33, 1073–1082.
- \*Holtzman, N. S. (2010). *The evolutionary persistence of socially aversive personality traits* (Unpublished doctoral dissertation). Washington University, St. Louis, MO.
- \*Holtzman, N. S., & Strube, M. J. (2009, February). *A reliable implicit association test of narcissistic personality disorder (the NPD-IAT)*. Poster session presented at the meeting of the Society for Personality and Social Psychology, Tampa, FL.
- \*Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: Behavioral manifestations of narcissism in everyday life. *Journal of Research in Personality*, 44, 478–484.
- \*Horton, R. S. (2011). [NPI-40, Rosenberg self-esteem, psychological health]. Unpublished raw data.
- \*Horton, R. S., Bleau, G., & Drwecki, B. (2006). Parenting narcissus: What are the links between parenting and narcissism? *Journal of Personality*, 74, 345–376.
- \*Horvath, S., & Morf, C. C. (2009). Narcissistic defensiveness: Hypervigilance and avoidance of worthlessness. *Journal of Experimental Social Psychology*, 45, 1252–1258.
- \*Horvath, S., & Morf, C. C. (2010). To be grandiose or not to be worthless: Different routes to self-enhancement for narcissism and self-esteem. *Journal of Research in Personality*, 44, 585–592.
- Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative knowledge in psychology. *International Journal of Selection and Assessment*, 8, 275–292.
- Hyer, S. E. (1994). *Personality Diagnostic Questionnaire-4*. Unpublished test, NYSPI.
- Johnson, J. A. (2004). The impact of item characteristics on item and scale validity. *Multivariate Behavioral Research*, 39, 273–302.
- \*Jordan, C. (2006). [NPI and Rosenberg self-esteem]. Unpublished raw data.
- \*Kansi, J. (2011). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Kansi, J. (2003). The Narcissistic Personality Inventory: Applicability in a Swedish population sample. *Scandinavian Journal of Psychology*, 44, 441–448.
- \*Konrath, S. (2004). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Konrath, S. (2005a). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Konrath, S. (2005b). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Konrath, S. (2007). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Konrath, S. (2010). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- Lakey, C. E., Rose, P., Campbell, W. K., & Goodie, A. S. (2008). Probing the link between narcissism and gambling: The mediating role of judgment and decision-making. *Journal of Behavioral Decision Making*, 21, 113–137.
- \*Locke, K. D. (2009). Aggression, narcissism, self-esteem, and the attribution of desirable and humanizing traits to self versus others. *Journal of Research in Personality*, 43, 99–102.
- Lynam, D., Hoyle, R., & Newman, J. (2006). The perils of partialling: Cautionary tales from aggression and psychopathy. *Assessment*, 13, 328–341.
- Lynam, D. R., & Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology*, 110, 401–412.
- \*Maples, J. L., Miller, J. D., Wilson, L. F., Seibert, L. A., Few, L. R., & Zeichner, A. (2010). Narcissistic personality disorder and self-esteem: An examination of differential relations with self-report and laboratory-based aggression. *Journal of Research in Personality*, 44, 559–563.
- \*Maxwell, K. M., Donnellan, M. B., Hopwood, C. J., & Ackerman, R. A. (2011). The two faces of narcissus? An empirical comparison of the Narcissistic Personality Inventory and the pathological narcissism inventory. *Personality and Individual Differences*, 50, 577–582.
- \*Miller, J. D., & Campbell, W. K. (2008). Comparing clinical and social-personality conceptualizations of narcissism. *Journal of Personality*, 76, 449–476.



- Miller, J. D., & Campbell, W. K. (2010). The case for using research on trait narcissism as a building block for understanding narcissistic personality disorder. *Personality Disorders: Theory, Research, and Treatment*, 3, 180–191.
- Miller, J. D., & Campbell, W. K. (2011). Addressing criticisms of the Narcissistic Personality Inventory (NPI). In W. K. Campbell & J. D. Miller (Eds.), *Handbook of narcissism and narcissistic personality disorder: Theoretical approaches, empirical findings, and treatments* (pp. 146–152). Hoboken, NJ: Wiley.
- \*Miller, J. D., Campbell, W. K., & Pilkonis, P. A. (2007). Narcissistic personality disorder: Relations with distress and functional impairment. *Comprehensive Psychiatry*, 48, 170–177.
- Miller, J. D., Campbell, W. K., Pilkonis, P. A., & Morse, J. Q. (2008). Assessment procedures for narcissistic personality disorder: A comparison of the Personality Diagnostic Questionnaire-4 and best-estimate clinical judgments. *Assessment*, 15, 483–492.
- Miller, J. D., Dir, A., Gentile, B., Wilson, L., Pryor, L. R., & Campbell, W. K. (2010). Searching for a vulnerable dark triad: Comparing factor 2 psychopathy, vulnerable narcissism, and borderline personality disorder. *Journal of Personality*, 78, 1529–1564.
- Miller, J. D., Gaughan, E. T., Pryor, L. R., Kamen, C., & Campbell, W. K. (2009). Is research using the Narcissistic Personality Inventory relevant for understanding narcissistic personality disorder? *Journal of Research in Personality*, 43, 482–488.
- Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Campbell, W. K. (in press). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality*.
- \*Miller, J. D., Maples, J., & Campbell, W. K. (this issue). Comparing the construct validity of scales derived from the Narcissistic Personality Inventory: A reply to Rosenthal and Hooley (2010). *Journal of Research in Personality*, 45, 401–407.
- Miller, J. D., Widiger, T. A., & Campbell, W. K. (2010). Narcissistic personality disorder and the DSM-5. *Journal of Abnormal Psychology*, 119, 640–649.
- \*Morf, C. C. (2010a). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Morf, C. C. (2010b). [NPI-40 and Rosenberg self-esteem]. Unpublished raw data.
- \*Morf, C. C. (2010c). [NPI-40, Rosenberg self-esteem, Buss-Perry aggression]. Unpublished raw data.
- \*Morf, C. C., Horvath, S., & Zimmermann, T. D. (submitted for publication). In the mind of a male narcissist: Worthlessness primes aggression.
- O'Shea, P. G., & Gustafson, S. B. (1999). *Confirmatory item response theory: An analysis of the Narcissistic Personality Inventory*. Poster presented at the 11th annual conference of the American Psychological Society, Denver, CO.
- Okada, R. (2009). Narcissism and psychological health in adolescence: A meta-analysis. *Japanese Journal of Developmental Psychology*, 20, 428–436.
- National Research Council (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy Press.
- Pappas, B. P., & O'Carroll, R. E. (1998). Extremes of self-esteem and narcissism and the experience and expression of anger and aggression. *Aggressive Behavior*, 24, 421–438.
- \*Park, S. W. (2008). *Personality and situational factors in judgment of responsibility* (Unpublished master's thesis). University of Dayton, Dayton, OH.
- \*Park, S. W., Bauer, J. J., & Ar buckle, N. B. (2009). Growth motivation attenuates the self-serving attribution. *Journal of Research in Personality*, 43, 914–917.
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait selfenhancement: A mixed blessing? *Journal of Personality and Social Psychology*, 74, 1197–1208.
- Paulhus, D. L. (2001). Normal narcissism: Two minimalist accounts. *Psychological Inquiry*, 12, 228–230.
- 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.
- \*Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the pathological narcissism inventory. *Psychological Assessment*, 21, 365–379.
- Pincus, A. L., & Lukowitsky, M. R. (2010). Pathological narcissism and narcissistic personality disorder. *Annual Review of Clinical Psychology*, 6, 421–446.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychology Measurement*, 1, 385–401.
- Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., et al. (2006). The Reactive-Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, 32, 159–171.
- Raskin, R., & Hall, C. S. (1979). A Narcissistic Personality Inventory. *Psychological Reports*, 45, 590.
- Raskin, R., & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternate form reliability and further evidence of construct validity. *Journal of Personality Assessment*, 45, 159–162.
- Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, 54, 890–902.
- Reidy, D. E., Foster, J. D., & Zeichner, A. (2010). Narcissism and unprovoked aggression. *Aggressive Behavior*, 36, 414–422.
- Rhodewalt, F., & Morf, C. C. (1998). On self-aggrandizement and anger: A temporal analysis of narcissism and affective reactions to success and failure. *Journal of Personality and Social Psychology*, 74, 672–685.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27, 151–161.
- Rose, P. (2002). The happy and unhappy faces of narcissism. *Personality and Individual Differences*, 33, 379–391.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.
- \*Rosenthal, S. A., & Hooley, J. M. (2010). Narcissism assessment in social-personality research: Does the association between narcissism and psychological health result from a confound with self-esteem? *Journal of Research in Personality*, 44, 453–465.
- Rosenthal, S. A., & Pittinsky, T. L. (2006). Narcissistic leadership. *The Leadership Quarterly*, 17, 617–633.
- \*Samuel, D. B. (2008). [NPI-40, Rosenberg self-esteem, psychological health]. Unpublished raw data.
- Samuel, D. B., & Widiger, T. A. (2004). Clinicians' personality descriptions of prototypic personality disorders. *Journal of Personality Disorders*, 18, 286–308.
- \*Samuel, D. B., & Widiger, T. A. (2008). Convergence of narcissism measures from the perspective of general personality functioning. *Assessment*, 15, 364–374.
- Sedikides, C., Rudich, E. A., Gregg, A. P., Kumashiro, M., & Rusbult, C. (2004). Are normal narcissists psychologically healthy? Self-esteem matters. *Journal of Personality and Social Psychology*, 87, 400–416.
- Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, 53, 80–93.
- \*Sutin, A. R., & Robins, R. W. (2005). Continuity and correlates of emotions and motives in self-defining memories. *Journal of Personality*, 73, 793–824.
- \*Sutin, A. R., & Robins, R. W. (2007). Phenomenology of autobiographical memories: The memory experiences questionnaire. *Memory*, 15, 390–411.
- Tellegen, A., & Waller, N. G. (1994). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In S. R. Briggs & J. M. Cheek (Eds.), *Personality measures: Development and evaluation* (Vol. 1, pp. 133–161). Greenwich, CT: JAI Press.
- Thomaes, S., Stegge, H., Bushman, B. J., Olthof, T., & Denissen, J. (2008). Development and validation of the Childhood Narcissism Scale. *Journal of Personality Assessment*, 90, 382–391.
- \*Tracy, J. L., Cheng, J. T., Robins, R. W., & Trzesniewski, K. H. (2009). Authentic and hubristic pride: The affective core of self-esteem and narcissism. *Self and Identity*, 8, 196–213.
- \*Troisi, J. D., & Park, L. (2006). [NPI-40, Rosenberg self-esteem, aggression, psychological health]. Unpublished raw data.
- \*Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2008). Do today's young people really think they are so extraordinary? *Psychological Science*, 19, 181–188.
- Twenge, J. M., & Campbell, W. K. (2003). "Isn't it fun to get the respect we're going to deserve?" Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin*, 29, 261–272.
- Twenge, J. M., Konrath, S., Foster, J. D., Campbell, W. K., & Bushman, B. J. (2008). Egos inflating over time: A cross-temporal meta-analysis of the Narcissistic Personality Inventory. *Journal of Personality*, 76, 875–901.
- \*University of California, Davis (2010). [Psychology Department Mass Testing: NPI-40, Rosenberg self-esteem, psychological health, 2009–2010]. Unpublished raw data.
- Wallace, H. M., & Baumeister, R. F. (2002). The performance of narcissists rises and falls with perceived opportunity for glory. *Journal of Personality and Social Psychology*, 82, 819–834.
- Wright, A. G. C., Lukowitsky, M. R., Pincus, A. L., & Conroy, D. E. (2010). The higher order factor structure and gender invariance of the Pathological Narcissism Inventory. *Assessment*, 17, 467–483.
- Yarkoni, T. (2010). The abbreviation of personality, or how to measure 200 personality scales with 200 items. *Journal of Research in Personality*, 44, 180–198.