Narcissism assessment in social–personality research: Does the association between narcissism and psychological health result from a confound with self-esteem?

Seth A. Rosenthal *, Jill M. Hooley

Harvard University, Department of Psychology, 33 Kirkland Street, Cambridge, MA 02138, United States

A R T I C L E  I N F O

Article history:
Available online 15 May 2010

Keywords:
Narcissism
Self-esteem
Psychological health
Narcissistic personality inventory (NPI)
Narcissistic personality disorder
Pathological narcissism
Normal narcissism

A B S T R A C T

Influential social and personality psychology research indicates that narcissism is related to psychological health. Such inferences are open to question, however, because they nearly all rely on the same self-report instrument—the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979, 1981)—to operationalize and measure narcissism. This is problematic because numerous NPI items do not appear to correspond to common definitions or manifestations of narcissism, and may instead be indicative of self-esteem. Two studies demonstrate that the NPI’s confound with self-esteem accounts for the purported relationship between narcissism and psychological health. This suggests that inferences about narcissism and psychological health may need to be reinterpreted. Results also highlight the need for measures that correspond more directly to core components of narcissism.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Imagine hearing about two people who are both described as assertive and confident. Both also prefer leadership roles and enjoy receiving compliments. However, beyond these apparent similarities, the two are quite different from each other. The first is consistently full of bravado to overcompensate for core insecurities, insatiably in need of other people’s admiration (e.g., Morf & Rholes, 2001), and unreasonably pushy, overbearing, and demanding. You probably would not be surprised to hear this person referred to as a narcissist (e.g., American Psychiatric Association, 2000). In contrast, the second person is “happier than most, less so-anxious, … less depressed, … [and] higher [in] self-esteem” (see W. K. Campbell, 2001, p. 214). Is it possible that this normatively happy and healthy person is a narcissist as well? According to some perspectives in social and personality psychology research (e.g., W. K. Campbell, 2001; Miller & Campbell, 2008), the answer may be “yes.” In fact, the psychologically healthy characteristics enumerated above are a partial description of a narcissist from one social–personality point of view.

The narcissist with these healthy traits lies at the center of a crucial question about the limits of the theoretical and operational definitions of narcissism. To what extent should high self-esteem, and normative traits related to it such as assertiveness and confidence, play a role in defining narcissism? The divergence between normative social–personality perspectives on this question, and clinical perspectives, which generally emphasize narcissists’ low or fragile self-esteem (e.g., APA, 2000; Kernberg, 1975; Kohut, 1971, 1977), has made it increasingly difficult to reconcile social–personality and clinical narcissism research (Cain, Pincus, & Ansell, 2008; Miller & Campbell, 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009). Moreover, and most germane to this article, disagreement over whether self-esteem and other normative traits should be considered narcissistic has created confusion about whether psychological health is truly a potential characteristic of people who are narcissistic.

Although it is clear that many social–personality psychologists include high self-esteem in their conceptions of narcissism whereas clinical psychologists generally do not (Fossati et al., 2005; Miller & Campbell, 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009), it is unlikely that this reflects fundamental theoretical differences between the two camps. Instead, we posit that the inclusion of high self-esteem and other normative characteristics in social–personality definitions of narcissism results primarily from the way narcissism is traditionally operationalized. The self-report scale that is used to measure narcissism in nearly all social–personality research (Cain et al., 2008; Mullins & Kopelman, 1988; Pincus et al., 2009) is the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979, 1981). The NPI contains items that appear to be confounded with self-esteem, a characteristic that they may measure at least as effectively as they measure narcissism. As a result, the NPI’s correlates include not only the negative psychological states and outcomes that might be shared with typical...
clinical definitions of narcissism (e.g., aggression, Bushman & Baumeister, 1998; reduced romantic commitment, W. K. Campbell & Foster, 2002; pathological gambling, Lakey, Rose, Campbell, & Goodie, 2008), but also the positive psychological states that are more likely to be associated with high self-esteem, such as those noted in the description of the happy, healthy narcissist in example two.

Self-esteem is the normative characteristic with which the NPI appears to share its most robust empirical relationship (see Brown & Zeigler-Hill, 2004; Cain et al., 2008; Raskin, Novacek, & Hogan, 1991a, 1991b; Rhodewalt & Morf, 1995). One reason to closely examine the NPI's positive link with self-esteem, and the possibility that this link is largely a byproduct of confounded items contained within the NPI itself, is that there is accumulating evidence that the connection between the NPI and psychological health is contingent on the NPI's relationship with self-esteem. This is the case, for example, in research by Rose (2002), who reported that narcissism was related to multiple measures of happiness, and by Sedikides, Rudich, Gregg, Kumashiro, and Rusblut (2004), who reported that narcissism was related to numerous indicators of psychological health, such as lower depression, sadness, anxiety, and neuroticism, as well as higher levels of personal and couple well-being. In both sets of studies, the relationship between the NPI and each of its healthy psychological correlates was fully mediated by self-esteem (see also Brown, Budzak, & Tamborski, 2009). This led Sedikides and colleagues to conclude that self-esteem is the "key component of narcissism" (p. 401) through which narcissism predicts positive psychological outcomes (see also Sinha & Krueger, 1998). In contrast, self-esteem does not appear to mediate the association between the NPI and the kinds of negative outcomes more traditionally associated with narcissism. For example, controlling for self-esteem did not negate the relationship between the NPI and reduced romantic commitment (W. K. Campbell & Foster, 2002), pathological gambling (Lakey et al., 2008), or the greedy exploitation of natural resources (W. K. Campbell, Bush, Brunell, & Shelton, 2005). Taken together, it appears that a link between the NPI and psychological health (but not negative psychological outcomes) depends on a theory of narcissism that considers high self-esteem to be an inherent narcissistic trait (e.g., Baumeister, Campbell, Krueger, & Vohs, 2003).

To justify the inclusion of self-esteem and other normative characteristics within a social–personality theory of narcissism, some have called for revisiting historic clinical theories that provide the grounds doing so (see W. K. Campbell, 2001; Miller & Campbell, 2008). However, such an approach may be problematic for several reasons. First, it employs post hoc theoretical explanations for findings that might be more parsimoniously explained through an empirical critique of some of the NPI's items. Further, the clinical theories that did include high self-esteem in their definition of narcissism were either too nebulous and overinclusive (e.g., Freud, 1931/1950) or too multifaceted (e.g., Kohut, 1971, 1977) to be operationalized properly using any single measure (see Clark & Watson, 1995; Raskin & Terry, 1988).

Perhaps most problematically, these historical clinical theories are not aligned with the theoretical foundation that was used to develop the NPI. The NPI was not developed to correspond with a normative definition of narcissism. Instead, the scale's authors intended to provide a continuous, non-clinical measure (for use in normal populations) of the characteristics described by a specific clinical definition of narcissism, the diagnostic criteria for narcissistic personality disorder (NPD) in the then-current edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-III, APA, 1980; see Raskin & Terry, 1988). Accordingly, the NPI's authors developed items that they believed "sampled the domain of the narcissistic personality" (Raskin & Hall, 1979, p. 590) identified by the DSM-III criteria. These criteria included a grandiose sense of self-importance, preoccupation with fantasies of unlimited success, exhibitionism, indifference or rage, and disturbances in interpersonal relationships such as entitlement, exploitative-ness, splitting, and lack of empathy. The NPI was based on the assumption that "abnormality is continuous with normality, [and thus] behaviors descriptive of the Narcissistic Personality Disorder are … extreme forms [of these narcissistic traits] which are manifested to a lesser extent in normal individuals" (Raskin & Hall, 1981, p. 159).

In other words, the scale's authors intended for the NPI to measure less extreme levels of characteristics associated with narcissism as described in the DSM-III (see Pincus & Lukowitsky, 2010; Pincus et al., 2009). They did not intend to expand the definition of the construct based on normative clinical or social–personality theories. By referring to historic theories that included high self-esteem in their definition of narcissism to justify NPI-based findings about psychological health, critics (e.g., W. K. Campbell, 2001; Miller & Campbell, 2008) implicitly suggest that the authors of the NPI might use the terms narcissism and self-esteem interchangeably. The NPI's authors appear to have created a scale that is not entirely consistent with the theory on which it was based. In fact, the NPD section of the DSM-III states that narcissistic self-esteem is "often fragile" (p. 316), rather than that Narcissists have high self-esteem.

Using the DSM-III criteria as a theoretical "conceptual template" (Raskin & Terry, 1988, p. 892), it is easy to understand how items ranging from severe manifestations of narcissistic characteristics (e.g., "I insist on getting the respect that is due me;") to more moderate manifestations (e.g., "I am apt to show off if I get a chance;" "I am more capable than other people") were included in the NPI. Unfortunately, other NPI items, such as "I am assertive" and "I see myself as a good leader" appear to fall outside of the NPI's authors' own DSM-III-based definition of narcissism. These more normative characteristics do not appear to lie on a clear continuum with those enumerated in the DSM-III. Rather, they most likely measure subjectively different constructs from the diagnostic criteria rather than less severe versions of the same constructs (see Pincus & Lukowitsky, 2010; Pincus et al., 2009).

A practical problem this causes is that the NPI's most normative items are not specific to narcissism—a narcissist might be described by them (if only superficially), as in the first example presented earlier. But the content of these items also might not be indicative of narcissism. In other words, the items do not necessarily differentiate well between narcissists and non-narcissists. Even if narcissists often exhibit these characteristics (see Lynam & Widiger, 2001; Miller, Gaughan, Pryor, Kamen, & Campbell, 2009; Samuel & Widiger, 2004), knowing that someone possesses these attributes does not necessarily help one discern how narcissistic that individual is, or even whether that individual is narcissistic at all. Further, such items may provide more information about whether a person has high self-esteem than about whether that person is narcissistic. And although such normative attributes are not measured by a majority of the NPI's items, the internal consistency strategy originally used to develop the NPI resulted in statistical overrepresentation of its most normative items in aggregated scores (Emmons, 1984; Kanski, 2003; Raskin & Terry, 1988). As a result, aggregated NPI scores are poorly proportioned (see Haynes, Richard, & Kubany, 1995), causing them to follow the scale's most normative items in correlating with variables indicative of psychological health.

In sum, the conclusion that narcissism is associated with healthy psychological states and outcomes may depend largely on the effects of a number of items in the NPI that appear to be more closely related to self-esteem than to narcissism. To investigate this possibility, in Study 1 we examined whether the NPI's...
overlap with self-esteem is attributable to a theoretically sound empirical relationship between the two constructs, or whether, as we hypothesize, it is the result of the inclusion of specific items in the NPI that are confounded with self-esteem. In Study 2, we explored whether these same NPI items account for the types of relationships between narcissism and psychological health that have been reported in the social–personality literature.

2. Study 1

2.1. Overview

In Study 1, we used three independent methods—expert ratings, item response theory (IRT) data, and traditional NPI factors derived using exploratory factor analysis (EFA)—to delineate which NPI items did a good job measuring narcissism, and which items functioned relatively poorly at measuring the construct. We then combined the items that most clearly measured narcissism into NPI-N subscales, whereas the items that appeared to measure some other characteristic(s) better than they measured narcissism were combined and labeled the NPI-X subscales. We then investigated the subscales’ relationships to a benchmark measure of DSM-based narcissism to determine whether the subscales measured narcissism in a manner consistent with the intentions of the NPI’s authors. We also used a benchmark measure of self-esteem to investigate whether the NPI-X subscales measured self-esteem more effectively than they measured DSM-based narcissism.

2.2. Method

2.2.1. Participants

The sample consisted of 232 students between the ages of 17 and 24–129 men and 103 women. The mean age of participants was 19.5 years (SD = 1.2). This sample is typical of the population with which NPI-based research is often conducted. Participants were compensated with research credit or small payments. They were recruited through a psychology department study pool and also via posters.

2.2.2. Procedure

Participants provided informed consent and demographic information. They then completed the narcissism and self-esteem questionnaires as part of a larger study. To reduce priming effects, the narcissism and self-esteem questionnaires were each separated by several other questionnaires that did not measure either construct. At the end of the study, participants were debriefed and informed about the purpose of the research.

2.2.3. Materials

2.2.3.1. Narcissistic Personality Inventory. The NPI is a forced-choice measure of narcissism as a dimensional a personality trait. It includes items ranging in severity from “I am assertive” to “I insist on getting the respect that is due me.” Participants completed the 37-item, 4-factor version of the scale (Emmons, 1987).

2.2.3.2. Personality Questionnaire for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders. Participants completed the NPD section of the Personality Questionnaire for the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (PQ; First, Gibbon, Spitzer, Williams, & Benjamin, 1997). This allowed us to test the convergent validity of NPI items against a benchmark measure directly related to the construct on which the NPI is purportedly based. The PQ is a 119-item yes/no questionnaire. Its items correspond to the DSM-IV diagnostic criteria for personality disorders, and are directly linked to the content of the Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II; First et al., 1997). Participants completed the 17 PQ items that correspond to the nine DSM-IV diagnostic criteria for NPD. Examples include “Do people often fail to appreciate your very special talents or accomplishments?” and “Do you often expect other people to do what you ask without question because of who you are?” Responses were coded according to the number of NPD criteria endorsed (i.e., scores ranged from 0 to 9). For criteria with more than one associated PQ item, a “yes” answer to any associated item was coded as endorsement of the criterion.

The PQ was designed to be a screening instrument for the SCID-II diagnostic interview (First, Spitzer, Gibbon, & Williams, 1995). However, agreement between the two assessments is such that the PQ can serve as a reasonable alternative to the interview (Ekslius, Lindstrom, von Knorring, Bodlund, & Kullgren, 1994). The PQ has been used previously as a stand-alone research instrument (e.g., Dowson & Berrios, 1991; Morse, Robins, & Gittes-Fox, 2002).

2.2.3.3. Rosenberg Self-Esteem Scale. The Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1965) is a widely used measure that assesses participants’ feelings of self-worth. It contains ten items such as “On the whole, I am satisfied with myself” rated on a 1 (strongly disagree) to 4 (strongly agree) scale. The RSE has been recommended for use in research that aims to separate the effects of narcissism from those of self-esteem (Brown & Zeigler-Hill, 2004).

2.3. Approaches to assessing the narcissistic content of NPI items

2.3.1. Expert ratings

Ratings of the narcissistic content of NPI items were made by 19 clinicians who were also empirical narcissism researchers. Potential participants met two criteria: they were (a) Ph.D. or M.D. level clinicians and (b) authors of empirical peer-reviewed narcissism research articles. Using the PsycINFO database, we identified 32 individuals who met both criteria. We contacted each potential participant by email and sent a follow up request if there was no response to the first message. Nineteen participants (16 men and three women) completed the study. Of these, 15 held Ph.D.s, three held M.D.s., and one held both degrees.

Experts assessed the content validity of each of the 54 items from the original version of the NPI (Raskin & Hall, 1979, 1981) by rating (a) whether it was “indicative of the construct of narcissism” and (b) whether it suggested a “clinical or diagnostic level of narcissism.” This second step was intended, in part, to remind participants that items did not have to meet a clinical threshold in order to be considered indicative of the construct. Expert raters were not directed to make their ratings based on any particular theoretical or typological orientation. Instead, they were intended to consider narcissism “from their own diverse experiences” (see Lynam & Widiger, 2001, p. 409).

Expert ratings and subscale information for the 37 items included in the Emmons (1987) version of the NPI are presented in Table 1. Of the 37, the mean number of items each rater regarded as indicative of the construct of narcissism (i.e., hits) was 20.55 (SD = 5.67). From the perspective of the scale, of the 703 total item ratings made, 390 (55.5%) were hits (SD = 0.16). From the perspective of items, the mean number of hits per item (out of a possible 19) was 10.54 (SD = 4.16). Overall, 25 of the 37 items were endorsed as indicative of the construct of narcissism by more than half of the experts. The two items that research has shown to have the strongest statistical relationship with aggregated NPI scores (i.e., they have the greatest statistical influence on overall NPI scores; Raskin & Terry, 1988) were the “I would prefer to be a leader” and “I like to have authority over people” received only two and ten hits respectively. In contrast, the item that has been shown to have the weakest relationship with the aggregated NPI (i.e., it has the
least statistical influence on overall NPI scores; Raskin & Terry, 1988), “I insist upon getting the respect that is due me,” received the greatest number of hits (17) from the expert raters.

Overall agreement among the 19 raters was good. The two-way mixed, absolute agreement, average measures intraclass correlation, which assumes that these specific experts’ ratings will be used in the aggregate, was .77. The alpha for raters’ internal consistency (with raters treated as variables and NPI items treated as cases) was .79. No single rater had a significantly detrimental effect on the alpha coefficient.

Using a liberal 50% endorsement threshold for each item as a cut-off (see Lawshe, 1975), we classified the 25 NPI items rated as narcissistic by 10 or more expert raters as Expert-N items, and the 12 items rated as narcissistic by nine or fewer raters as Expert-X items.

2.3.2. Item response theory

In the context of the NPI, an IRT analysis can determine whether each NPI item is likely to be endorsed mainly by individuals who were higher in the general trait measured by the scale, or is just as likely to be endorsed by individuals with low levels of the trait. O’Shea and Gustafson (1999) conducted a three-parameter model IRT analysis on the 40-item version of the NPI (Raskin & Terry, 1988) with data from 1260 undergraduate participants. The 40-item NPI shares 31 items in common with the Emmons (1987) 37-item version employed in the current research. O’Shea and Gustafson’s (1999) results indicated that a number of the scale’s items measured traits toward the lowest severity levels of the NPI’s spectrum. As indicated by b parameter scores near or below zero, these items were likely to be endorsed in the “narcissistic” direction by individuals who had near average, or even below average levels of the general construct measured by the NPI. The data identified 13 items with low b parameter scores. The authors also identified a 14th item with a high c parameter score, which indicated that it was excessively likely to be endorsed by individuals who were extremely low in the trait measured by the scale. Thirteen of the 14 identified items are also contained in the Emmons (1987) 37-item NPI. Accordingly, we classified those 13 items as IRT-X items. The other eighteen items shared by both scales (i.e., those with higher b parameter scores) were classified as IRT-N items. We dropped the remaining six items for which there were no ratings.

2.3.3. Traditional subscales

Emmons (1984, 1987) derived four NPI subscales using EFA: Exploitativeness/Entitlement (E/E), Superiority/Arrogance (S/A), Self-Absorption/Self-Sufficiency (S/S) subscale, X = item from the Leadership/Authority (L/A) subscale.

Table 1
Expert, IRT, and EFA subscale ratings for 37-item NPI, Study 1.

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert hits</th>
<th>Expert subscale</th>
<th>IRT subscale</th>
<th>EFA subscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. I insist upon getting the respect that is due me</td>
<td>17</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>35. I will never be satisfied until I get all that I deserve</td>
<td>16</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14. I think I am a special person</td>
<td>16</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>54. I am an extraordinary person</td>
<td>16</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>48. I am going to be a great person</td>
<td>16</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>32. Everybody likes to hear my stories</td>
<td>15</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>4. Superiority is something you are born with</td>
<td>14</td>
<td>N</td>
<td>–</td>
<td>N</td>
</tr>
<tr>
<td>49. I can make anybody believe anything I want them to</td>
<td>14</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>7. I know that I am good because everybody keeps telling me so</td>
<td>14</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>18. People can learn a great deal from me</td>
<td>14</td>
<td>N</td>
<td>–</td>
<td>N</td>
</tr>
<tr>
<td>40. I am envious of other people’s good fortunes</td>
<td>13</td>
<td>N</td>
<td>–</td>
<td>N</td>
</tr>
<tr>
<td>46. People always seem to recognize my authority</td>
<td>13</td>
<td>N</td>
<td>N</td>
<td>X</td>
</tr>
<tr>
<td>29. I always know what I am doing</td>
<td>13</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>19. I find it easy to manipulate people</td>
<td>12</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>53. I am more capable than other people</td>
<td>11</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>26. I like to look at my body</td>
<td>11</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>28. I am apt to show off if I get a chance</td>
<td>11</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>33. I usually dominate any conversation</td>
<td>11</td>
<td>N</td>
<td>–</td>
<td>N</td>
</tr>
<tr>
<td>42. I like to look at myself in the mirror</td>
<td>11</td>
<td>N</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>38. I have a strong desire for power</td>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>52. I get upset when people don’t notice how I look when I go out in public</td>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10. I can usually talk my way out of anything</td>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>22. I can read people like a book</td>
<td>10</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2. I have a natural talent for influencing people</td>
<td>10</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>17. I like having authority over other people</td>
<td>10</td>
<td>N</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>36. I like to be complimented</td>
<td>9</td>
<td>X</td>
<td>X</td>
<td>N</td>
</tr>
<tr>
<td>12. I like to be the center of attention</td>
<td>9</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>44. I really like to be the center of attention</td>
<td>9</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>50. I am a born leader</td>
<td>9</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>21. I like to display my body</td>
<td>9</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>27. I have good taste when it comes to beauty</td>
<td>7</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6. I would be willing to describe myself as a strong personality</td>
<td>6</td>
<td>X</td>
<td>–</td>
<td>X</td>
</tr>
<tr>
<td>15. I see myself as a good leader</td>
<td>6</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>34. I expect a great deal from other people</td>
<td>5</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>47. I would prefer to be a leader</td>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>16. I am assertive</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5. I do almost anything on a dare</td>
<td>0</td>
<td>X</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Mean (SD) 10.54 (4.16)

Note. For expert ratings, N = 19. Expert hits = number of experts who endorsed the item as indicative of narcissism. IRT = item response theory. EFA = exploratory factor analysis. Item numbers are from the original 54-item NPI. For the Expert subscales, N = more than 50% of raters endorsed item as indicative of narcissism, X = fewer than 50% of raters endorsed item as indicative of narcissism. For the IRT subscales, N = not identified as a poor item by IRT data, X = identified as a poor item by IRT data, – = IRT data not available. For the EFA subscales, N = item from the Exploitativeness/Entitlement (E/E), Superiority/Arrogance (S/A), or Self-Absorption/Self-Sufficiency (S/S) subscale, X = item from the Leadership/Authority (L/A) subscale.
Pincus, 2003; P. J. Watson, Grisham, Trotter, & Biderman, 1984). In contrast, L/A was considered the most adaptive and least clinically-relevant. The other two subscales, S/A and S/S, fell between E/E and L/A, but are often considered to be adaptive subscales that are closely related to L/A (Emmons, 1984, 1987; P. J. Watson & Biderman, 1993; P. J. Watson, Little, Sawrie, & Biderman, 1992; P. J. Watson & Morris, 1990).

However, there is also evidence that L/A stands apart from the two other adaptive subscales, S/A and S/S, as well as from E/E. L/A exhibits a unique lack of convergence with various other measures of narcissism (Shulman & Ferguson, 1988). It is also more strongly related to normative characteristics such as assertiveness (P. J. Watson, McKinney, Hawkins, & Morris, 1988) and extraversion (Emmons, 1984) than are any of the other subscales. Further, the L/A subscale is lacking in construct validity on its face (Emmons, 1987; Mullins & Kopelman, 1988), because leadership and related characteristics do not appear among the DSM-III (or subsequent) NPD criteria, but have long been measured as desirable non-narcissistic traits (e.g., Gough, 1969). Thus, the L/A subscale essentially “adds a new dimension” to narcissism (Mullins & Kopelman, 1988, p. 620), which is generally considered poor practice in defining the limits of a construct in order to measure it precisely (Clark & Watson, 1995; Haynes et al., 1995).

Because of the potential problems specific to the L/A subscale, we investigated the consequences of separating L/A from the rest of the NPI by labeling the nine L/A items the EFA-X subscale, and the 28 items from the other three subscales the EFA-N subscale. Thus, the EFA-N subscale retained the items from the adaptive S/A and S/S subscales as well as from the maladaptive E/E subscale.

3. Results

3.1. Aggregated NPI

Descriptive statistics, reliabilities, and correlations for the scales and subscales used in Study 1 are reported in Table 2. The aggregated NPI was significantly correlated with both the PQ and the RSE. Further, there was no statistical difference between the magnitudes of the two correlation coefficients (rs = .34 and .36 respectively; Z = 0.22, p = .83; see Meng, Rosenthal, & Rubin, 1992, for the method used to compare correlated correlation coefficients throughout this article). The same coefficients, corrected for attenuation due to less than perfect internal consistency (Spearman, 1904), were also statistically equivalent (disattenuated rs = .48 and .43 respectively, Z = 0.60, p = .53). In other words, when used as an aggregated scale, the NPI was as closely related to a benchmark measure of self-esteem as it was to a benchmark measure of narcissism. This is also notable as the two benchmark measures were negatively correlated with each other in the same data set (r = -.21, p < .01).

3.2. Subscales

We used three independent methods for dividing the NPI into two sets of items that were better, or more poorly, aligned with the construct of narcissism: clinical research experts’ content validity ratings, IRT results reflecting internal validity, and traditional EFA-derived subscales developed to investigate construct validity. As indicated in Table 1, the three methods produced partially overlapping results. For the 31 items rated by all three methods, there was three-way agreement on 18 (13 agreed as NPI-subscale), and disagreement on the 13 remaining items. Four of the six items for which there were only two ratings were agreed as NPI-N items, one as an NPI-X item, and there was disagreement on one.

Although there was notable variation in the composition of the subscales produced by the three methods for dividing the NPI, the pattern of the correlations between the subscales and the benchmark scales was nearly identical (see Table 2). In each case, the NPI-N subscale was significantly correlated with the benchmark measure of narcissism (PQ) whereas the NPI-X subscale was uncorrelated with the PQ. In contrast, all subscales were significantly correlated with the self-esteem benchmark measure (RSE). However, in each pair of subscales, the NPI-X subscale’s correlation with the RSE was significantly stronger than was the corresponding NPI-N subscale’s correlation (Z_{Expert} = 4.40; Z_{IRT} = 7.06; Z_{EFA} = 4.26, all ps < .001). This indicates that the NPI-X subscales share a stronger relationship with self-esteem than do the NPI-N subscales. More importantly, each NPI-X subscale was also more strongly correlated with the RSE than with the PQ (Z_{Expert} = 4.10; Z_{IRT} = 4.82; Z_{EFA} = 4.77, all ps < .001). This indicates that each NPI-X subscale was more strongly related to the benchmark measure of self-esteem than to the benchmark measure of narcissism.

We also investigated whether the NPI-X subscales contributed variance independent of the NPI-N subscales to predicting the benchmark narcissism measure by simultaneously entering the corresponding paired subscales in regressions predicting the PQ.
The left half of Table 3 illustrates that in each case, the NPI-N subscale predicted PQ scores. In contrast, each NPI-X subscale was negatively related to the PQ. Although this occurred in the context of statistical suppressor effects, whereby estimates of the specific direction and strength of the relationships are unreliable, the suppression did not affect the key conclusion that the NPI-X subscales were not positively related to the PQ.\footnote{The statistical suppressor effects do not appear to be the result of multicollinearity between the NPI-N and NPI-X subscales. Allison (1999) provided the conservative estimate that a variance inflation factor (VIF) of 2.50 or higher indicates a multicollinearity problem. In contrast, the VIF in each of the current regression models was lower than 1.40.}

In contrast, as illustrated in the right half of Table 3, when entered simultaneously in a regression, each of the NPI-X subscales predicted the benchmark self-esteem measure whereas the corresponding NPI-N subscales did not contribute any additional variance to predicting self-esteem (except in the case of the IRT-derived subscales, where IRT-N was negatively related to the RSE). This provides further evidence that the items comprising the NPI-X subscales are robustly related to self-esteem.

Finally, it is also noteworthy that the potential problems with the L/A subscale identified by the traditional subscale approach may have been recognized implicitly in the experts’ ratings. Only three of L/A’s nine items were rated as indicative of narcissism by more than half of the experts. The mean number of hits per item for the 9-item L/A subscale was 7.33 (SD = 3.94), and the mean hit percentage per item was 38.6% (SD = 0.20). In contrast, the hits per item for the 11-item S/A subscale was 11.00 (SD = 4.17), hit percentage = 57.9% (SD = 0.21); for the 9-item S/S subscale, hits per item = 12.11 (SD = 3.48), hit percentage = 63.2% (SD = 0.26); and for the 8-item Exploitativeness/Entitlement subscale (E/E), hits per item = 11.75 (SD = 3.77), hit percentage = 61.8% (SD = 0.24). Overall, the average item from the L/A subscale was endorsed as narcissistic significantly less frequently than was the average item from the three other subscales, paired-samples t(18) = 4.07, p < .001.

4. Discussion

Results from Study 1 support the hypothesis that portions of the NPI do not do a good job of measuring narcissism. Further, these items appear to be confounded with self-esteem. These conclusions are borne out almost identically whether the NPI is divided into subscales using judgments of expert clinical researchers, IRT analyses (O’Shea & Gustafson, 1999), or subscales based on EFAs conducted by a personality psychologist (Emmons, 1987).

The conclusions derive from two key findings. First, using each method to divide the NPI, the resulting NPI-X subscale is unrelated to the benchmark narcissism measure (i.e., the PQ). This suggests that major portions of the NPI do not meet the scale’s authors’ stated goal of measuring narcissism that is linked to the DSM diagnostic criteria for NPD (Raskin & Hall, 1979). Second, these less narcissistic NPI subscales have a significantly stronger relationship with the benchmark self-esteem measure (i.e., the RSE) than with the benchmark narcissism measure, suggesting that they measure self-esteem better than they measure narcissism.

The results of Study 1 also provide some potentially good news for the validity of inferences based on some of the NPI’s items. In all cases, the NPI-N subscales were strongly related to the benchmark narcissism scale. This indicates that many NPI items do meet the goal of measuring narcissism that overlaps with the DSM description of the trait (see Miller et al., 2009). The NPI-N subscales also shared positive zero-order relationships with the benchmark self-esteem scale, suggesting that the NPI’s more narcissistic items may also be positively related to self-esteem. However, the simultaneous regressions reported in Table 3 indicate that the relationships between the NPI-N subscales and self-esteem are negated by controlling for the NPI-X subscales, suggesting that this relationship is tenuous and in need of further investigation.

4.1. Limitations

It is important to note that each of the three methods used to divide the NPI into more and less narcissistic items poses problems on its own. An expert ratings approach is generally favored for test items that measure observable outcomes, such as job or school performance, rather than abstractions such as personality traits (Lawshe, 1985). Further, a 50% cut point for dividing the scale is arbitrary, although it is widely used and based on the principle that items with more than 50% endorsement by a group of experts have at least “some degree of content validity” (Lawshe, 1975, p. 567). Our design for assessing expert ratings also did not account for participants’ familiarity with or biases about the NPI, either of which could have affected their ratings. We also did not ask experts to choose whether each NPI item measures “narcissism or self-esteem,” which would have been more in keeping with the hypotheses of this research. Future research using expert ratings could benefit by including items from multiple narcissism and self-esteem measures, and providing a format in which experts choose whether each item measures narcissism or self-esteem.

IRT analyses assume that the measure being analyzed is unidimensional. However, the NPI has been shown in numerous studies to be multidimensional (Corry, Merritt, Mrug, & Pamp, 2008; Emmons, 1984, 1987; Kubarych, Deary, & Austin, 2004; Raskin & Terry, 1988). The IRT’s authors (O’Shea & Gustafson, 1999) agree that the NPI is not fully described by a single factor. On the other hand, they note that Monte Carlo simulations indicate that IRT models should produce accurate descriptions of a multidimensional scale if the scale’s first factor accounts for 20% or more of the overall variance (Reckase, 1979; see also Drasgow & Parsons, 1983). Accordingly, O’Shea and Gustafson (1999) conducted an EFA with their data. Their analysis produced a first factor that accounted for 43% of the variance in the NPI. This suggests that an IRT model of their data should produce relatively accurate results. Also, treating the NPI as if it were unidimensional introduces ecological validity to the research (if only unintentionally). Research using the NPI most often relies on a unidimensional aggregated score (Miller & Campbell, 2008). Further, the NPI was originally developed using an internal consistency strategy better suited for selecting scale items to measure a unidimensional than a multidimensional construct (Raskin & Terry, 1988). Nevertheless, because of dimensionality concerns, the IRT results should be treated with caution.

As noted earlier, Emmons’s (1987) traditional NPI subscales have typically been divided into three adaptive subscales (L/A, S/S, and S/A), with E/E set aside as the lone maladaptive subscale (Emmons, 1984, 1987; P. J. Watson et al., 1992). We take a different approach by combining E/E, S/S, and S/A into a single measure of narcissism and setting aside L/A as a less appropriate means for measuring narcissism. We believe that setting L/A, rather than E/E, aside is more theoretically consistent with the intentions of the NPI’s authors (Raskin & Hall, 1979, 1981). By removing L/A under the assumption that it is comprised of poorly-conceived items (i.e., that it does not measure narcissism properly), we retain the remainder of the NPI as a continuous measure of narcissism. In contrast, the common practice of separating E/E from the rest of the NPI while maintaining that both represent valid measures of narcissism suggests that the NPI is actually made up of two separate continuous measures of narcissism (E/E being a clinical-level narcissism measure that is discontinuous from the other three normal narcissism subscales; e.g., Dickenson & Pincus, 2003;
P. J. Watson et al., 1984, 1992). We believe that an approach to the NPI that retains the continuity from more to less severe narcissism, while removing the items that may not be closely associated with narcissism at all, is preferable. However, this method for dividing the NPI has not been tested empirically elsewhere, and should be considered with that in mind.

It is also important to note that the DSM diagnostic criteria for NPD have changed since the development of the NPI. This could account for some of our results (i.e., for the lack of a relationship between some NPI items and the PQ). However, we believe this is an unlikely explanation for a number of reasons. First, the crucial characteristics of grandiosity, entitlement, exploitativeness, need for admiration, lack of empathy, etc., were retained from the DSM-III, on which the NPI is based, through the DSM-IV, on which the PQ is based. Second, the changes in criteria from the inception of the DSM-III to the DSM-IV should theoretically increase, rather than decrease, the relationship between the NPI and the PQ. This is because the DSM criteria that were removed emphasized vulnerable narcissistic traits (e.g., inferiority, shame, emptiness, splitting; Cain et al., 2008; Gunderson, Ronningstam, & Smith, 1995). As a result, the remaining DSM criteria focus nearly exclusively on grandiose narcissistic themes (Cain et al., 2008; Miller & Campbell, 2008; Miller, Hoffman, Campbell, & Pilkonis, 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009), suggesting that there should be an increased level of overlap with the NPI. Accordingly, the changes in the diagnostic criteria for NPD made from the DSM-III to the DSM-IV are unlikely to be the cause of a lack of correlation between the NPI-X subscales and the PQ.

Overall, and even with these caveats, the results of Study 1 appear to support the central hypotheses: (1) that parts of the NPI do poor job of measuring narcissism and (2) that those poor parts of the NPI are confounded with and bias the aggregated scale toward measuring self-esteem.

5. Study 2

5.1. Overview

In Study 2, we investigated the degree to which the overlap between the NPI-X subscales and self-esteem was responsible for the relationship between the NPI and self-reports of indicators of psychological health and distress, including optimism, happiness, positive affect, aggression, anxiety sensitivity, and negative affect. We predicted that any relationship between the NPI and psychological health would be fully accounted for by the NPI-X subscales and by the NPI's overlap with self-esteem. In contrast, the literature reviewed earlier suggests that any connections between the NPI and psychological distress should be driven directly by the NPI-N subscales and should be minimally affected by controlling for self-esteem.

5.2. Method

5.2.1. Participants

A new sample of 141 students between the ages of 18 and 23–40 men and 101 women—was recruited through a psychology department study pool and also via recruitment posters. The mean age of participants was 19.6 years (SD = 1.1). This sample is typical of the population with which NPI-based research is often conducted. Participants were compensated with research credit or small payments.

5.2.2. Procedure

Participants provided informed consent and demographic information, and completed narcissism, self-esteem, and self-report psychological health and distress questionnaires as part of a larger study. Participants were then debriefed and informed about the study's purpose.

5.2.3. Materials

Participants completed the 37-item NPI and the RSE as well as self-reports of the following psychological health variables (PHVs):

- Optimism. The Life Orientation Test (Scheier & Carver, 1985) measures dispositional optimism. It consists of eight items such as “I always look on the bright side of things” rated on a 1 (strongly disagree) to 5 (strongly agree) scale.

- Happiness. The Oxford Happiness Questionnaire (Hills & Argyle, 2002) measures dispositional happiness. It consists of 29 items such as “I am very happy” rated on a 1 (strongly disagree) to 6 (strongly agree) scale.

- Positive and negative affect. The Positive and Negative Affect Schedule measures positive affect (i.e., “the extent to which a person feels enthusiastic, active, and alert;” D. Watson, Clark, & Tellegen, 1988, p. 1063) and negative affect (i.e., “subjective distress and unpleasurable engagement;” p. 1063). Each subscale consists of ten items which are rated on a 1 (very slightly or not at all) to 5 (extremely) scale. When used to measure dispositional affect, the Positive Affect and Negative Affect subscales are generally uncorrelated (Schmukle, Egloff, & Burns, 2002).

- Anxiety sensitivity. The Anxiety Sensitivity Index (Reiss, Peterson, Gursky, & McNally, 1986) measures the belief that anxiety experiences have negative implications and consequences (i.e., it measures individuals' fear of anxiety rather than their level of anxiety; Taylor, Koch, McNally, & Crockett, 1992). The scale contains 16 items (e.g., “It scares me when I am nervous”) which are rated on 1 (very little) to 5 (very much) scale. Anxiety sensitivity is associated with behavioral manifestations of anxiety (Maller & Reiss, 1987) and the presence of anxiety disorders, particularly panic disorder (McNally, 2002; Olatunji & Wolitzky-Taylor, 2009). It also shares a robust positive relationship with neuroticism (Norton, Cox, Hewitt, & McLeod, 1997; Zvolensky et al., 2003).

6. Results

Descriptive statistics for the scales and subscales from Study 2 are reported in Table 4. The first column of Table 5 reports the correlations between the NPI and the self-report PHVs. The aggregated NPI was positively related to each of the positive PHVs: optimism, happiness, and positive affect. The aggregated NPI was also positively related to aggression, but had a marginal negative relationship with the other two negative PHVs: negative affect and anxiety sensitivity. These are consistent with the relationships between narcissism and psychological health and distress typically reported in the types of social–personality studies reviewed earlier.

The second column reports the correlations between the RSE and the PHVs. The RSE was positively related to the each of the positive PHVs and negatively related to each of the negative PHVs. The aggregated NPI and RSE shared the same directional relationship with each of the PHVs except for aggression, with which the NPI was positively related whereas the RSE was negatively related.
Table 4
Descriptive statistics, Study 2.

<table>
<thead>
<tr>
<th>M</th>
<th>SD</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPI</td>
<td>13.54</td>
<td>6.19</td>
</tr>
<tr>
<td>Expert-N</td>
<td>7.58</td>
<td>3.72</td>
</tr>
<tr>
<td>Expert-X</td>
<td>5.34</td>
<td>2.86</td>
</tr>
<tr>
<td>IRT-N</td>
<td>4.75</td>
<td>2.98</td>
</tr>
<tr>
<td>IRT-X</td>
<td>7.13</td>
<td>3.31</td>
</tr>
<tr>
<td>EFA-N</td>
<td>8.44</td>
<td>4.33</td>
</tr>
<tr>
<td>EFA-X</td>
<td>5.10</td>
<td>2.82</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>3.19</td>
<td>0.54</td>
</tr>
<tr>
<td>Optimism</td>
<td>3.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Happiness</td>
<td>4.27</td>
<td>0.69</td>
</tr>
<tr>
<td>Positive affect</td>
<td>3.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Aggression</td>
<td>2.07</td>
<td>0.49</td>
</tr>
<tr>
<td>Anxiety sensitivity</td>
<td>2.22</td>
<td>0.57</td>
</tr>
<tr>
<td>Negative affect</td>
<td>2.06</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Note. N = 141. NPI = Narcissistic Personality Inventory. Expert = expert item content ratings. IRT = item response theory. EFA = exploratory factor analysis. Expert-, IRT-, and EFA-N subscales = subscales comprised of more clearly narcissistic NPI items. Expert-, IRT-, and EFA-X subscales = subscales comprised of less clearly narcissistic NPI items.

In each directionally-consistent case, the correlation between the RSE and the PHV was significantly stronger than the correlation between the aggregated NPI and the same PHV at the p < .01 significance level or greater.

Columns three through five report the correlations between the NPI-X subscales derived in Study 1 and the PHVs. In each case except for aggression, the NPI-X subscales shared the same directional relationship with the PHV as did both the aggregated NPI and the RSE. Each of the NPI-X subscales was uncorrelated with aggression—these relationships essentially “split the difference” between the positive relationship of the aggregated NPI with aggression and the negative relationship of the RSE with aggression. In every other case, the correlation between the RSE and the PHV was stronger than the correlation between the NPI-X subscale and the same PHV at the p < .05 level or greater, except for the RSE’s and NPI-X subscales’ correlations with positive affect, which were each statistically equivalent.

Columns six through eight report the correlations between the NPI-N subscales and the PHVs. The NPI-N subscales were significantly related to aggression and significantly or marginally related to positive affect. None of the other relationships between the NPI-N subscales and the PHVs were significant or marginal. In each case, the NPI-X subscale was more strongly correlated than its corresponding NPI-N subscale with each PHV in the direction that suggests a relationship between narcissism and psychological health at the p < .01 level or greater, except in the case of aggression, with which the NPI-N subscales were more strongly correlated than the NPI-X subscales in the “unhealthy” direction. In other words, whereas the NPI-X subscales, the aggregated NPI, and the RSE generally shared directionally consistent relationships with the PHVs, the NPI-N subscales’ relationships with the PHVs were generally inconsistent with those relationships. Accordingly, it appears that the NPI items that were identified in Study 1 as poorer indicators of the narcissism construct were largely responsible for the aggregated NPI’s relationship with the PHVs. In contrast, the NPI items that were deemed better indicators of narcissism in Study 1 appear to have little effect on the aggregated NPI’s relationship with the PHVs.

The right-most column reports partial correlations, which removed the variance associated with the RSE, between the aggregated NPI and the PHVs. Removing self-esteem’s variance from the aggregated NPI negated most of the significant relationships between the NPI and the PHVs, leaving a only a significant relationship with aggression and a marginal relationship with positive affect, and reversing (i.e., making positive) a marginal relationship with negative affect. Overall, the pattern of partial correlations between the aggregated NPI and the PHVs was similar to the pattern of zero-order correlations between the NPI-N subscales and the same PHVs. In contrast, this pattern was strikingly dissimilar from the pattern of zero-order correlations of the aggregated NPI with the PHVs.

7. Discussion

Investigating the connection of narcissism to psychological health using an aggregated NPI score paints a fairly straightforward picture. High NPI-rated narcissists appear to be more psychologically healthy than non-narcissists (e.g., W. K. Campbell, 2001). They are optimistic and happy, and are at least marginally lower in negative affect and fearful reactions to anxiety. In fact, they are healthy in ways very much like people with high self-esteem (although less strongly so), except that they report being aggressive whereas their high self-esteem counterparts report the opposite.

However, a closer examination of why there is a positive relationship between the aggregated NPI and self-reported psychological health reveals a starkly different picture. It is apparent that the relationship between the NPI and psychological health is nearly entirely dependent on a small number of NPI items—the same items that were deemed by multiple methods in Study 1 to be poor indicators of the narcissism construct. Study 2 further indicates that the NPI’s relationship with psychological health is largely a result of its overlap with self-esteem. Removing self-esteem from the aggregated NPI greatly reduced the scale’s relevance to the PHVs. This indicates that self-esteem plays a crucial role in guiding the NPI’s associations with self-reported psychological health.

In contrast, as predicted, the NPI’s association with one less normative trait—aggression—was largely driven by the NPI items that were deemed the better indicators of narcissism in Study 1. This relationship was actually strengthened by removing the effects of

Table 5

<table>
<thead>
<tr>
<th>NPI</th>
<th>Self-esteem</th>
<th>NPI-X subscales</th>
<th>NPI-N subscales</th>
<th>Partial correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expert</td>
<td>IRT</td>
<td>EFA</td>
<td>Expert</td>
</tr>
<tr>
<td>Optimism</td>
<td>.30</td>
<td>.70</td>
<td>.47</td>
<td>.51</td>
</tr>
<tr>
<td>Happiness</td>
<td>.29</td>
<td>.73</td>
<td>.49</td>
<td>.47</td>
</tr>
<tr>
<td>Positive affect</td>
<td>.35</td>
<td>.57</td>
<td>.48</td>
<td>.49</td>
</tr>
<tr>
<td>Aggression</td>
<td>.25</td>
<td>.36</td>
<td>.08</td>
<td>.03</td>
</tr>
<tr>
<td>Anxiety sensitivity</td>
<td>-.14</td>
<td>-.43</td>
<td>-.28</td>
<td>-.27</td>
</tr>
<tr>
<td>Negative affect</td>
<td>-.15</td>
<td>-.27</td>
<td>-.31</td>
<td>-.30</td>
</tr>
</tbody>
</table>

Note. N = 141. NPI = Narcissistic Personality Inventory. Expert = expert item content ratings. IRT = item response theory. EFA = exploratory factor analysis. Expert-, IRT-, and EFA-X subscales = subscales comprised of less clearly narcissistic NPI items. Expert-, IRT-, and EFA-N subscales = subscales comprised of more clearly narcissistic NPI items. Partial correlations (pr) remove the variance associated with self-esteem. All correlations r ≥ 14, p < .10; r ≥ .17, p < .05; r ≥ .22, p < .01; r ≥ .28, p < .001.
self-esteem (see Paulhus, Robins, Trzesniewski, & Tracy, 2004; Tracy, Cheng, Robins, & Trzesniewski, 2009).

Overall, the results of Study 2 indicate that either eliminating a small number of questionable items from the NPI, or controlling for the NPI's overlap with self-esteem, could potentially lead to over-hauling current findings in some social–personality literature about the connection of narcissism to psychological health. The ensuing results would likely indicate that such relationships are tenuous at best, and should be reconsidered.

8. General discussion

Much of the NPI-based research linking narcissism to psychological health may need to be reintegrated. It appears that establishing a robust connection between narcissism and psychological health necessitates the following: (1) using only the NPI (including its most normative items) to measure narcissism, (2) accepting that the NPI items that are most normative actually measure narcissism, and (3) accepting that the full mediation of the NPI's connection to psychological health by self-esteem reflects an inherent overlap of narcissism with high self-esteem rather than a confound.

Making each of these three choices is problematic. (1) Using only the NPI to measure narcissism is unwise both because relying on a single method of measuring any construct is generally considered problematic (D. T. Campbell & Fiske, 1959), and because, as our results indicate, aggregating scale items that may measure more than one construct can lead to making overinclusive inferences (Clark & Watson, 1995). (2) Accepting that the most normative NPI items measure narcissism is problematic because our results (as well as much of the accumulated literature) suggest that these items may bear little resemblance in their content or relationships with third variables to the DSM-III description of narcissism on which the NPI was purportedly based. Finally, (3) accepting that the NPI's overlap with self-esteem reflects a valid, robust relationship is problematic because many theoretical (e.g., Westen, 1990) and empirical (e.g., Brown & Bosson, 2001; Paulhus, 1998, 2001; Paulhus et al., 2004; Tracy et al., 2009) accounts of narcissism, most empirical work using narcissism measures other than the NPI (e.g., Soyer, Rovenpor, Kopelman, Mullins, & Watson, 2001), and our regression results, suggest that it does not.

If as a response to these problems, social–personality narcissism researchers modified their practice of relying on aggregated NPI scores, omitted questionable NPI items, and/or partialed out self-esteem as a confound when using the NPI, what we think we know about the positive side of narcissism could be radically different. Rather than appearing to be a sign of psychological health, narcissism from a social–personality perspective might align more closely with clinical perspectives on the trait.

8.1. Multiple definitions of narcissism?

Our results suggest that the NPI is most accurately described as measuring both less extreme levels of characteristics associated with clinical narcissism and self-esteem (and confounding the two into a single score) rather than as measuring a unitary construct of normal narcissism (see Miller & Campbell, 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009). However, it is true that, in contrast with most current clinical conceptions of narcissism, in narcissism's early theoretical incarnations, psychoanalysts such as Freud (1931/1950) and Rank (1911, cited in Pulver, 1970) also used the term "narcissism" to describe an eclectic variety of normative positive feelings and beliefs about the self analogous to high self-esteem (see W. K. Campbell, 2001; Miller & Campbell, 2008).

For instance, Freud (1931/1950) noted that narcissists can exhibit healthy unflappable strength and confidence, and suggested that any type of psychological investment in the self, whether pathological or not, fits under the umbrella of narcissism (see Pulver, 1970). Freud's clinical theory may align best with current social–personality theories that consider pathological and normal narcissism to be correlated constructs that are related through basic core traits (e.g., Miller & Campbell, 2008; Miller et al., 2009). Beyond those core traits, however, these types of narcissism have "substantially different nomological networks" (Miller & Campbell, 2008, p. 470), particularly in their relationship with psychological health. Kohut (1966) also differentiated normal from pathological narcissism, indicating that the two have independent developmental trajectories, where normal narcissism can develop into mature psychological structures and processes. His view may align best with current theories that consider pathological and normal narcissism to be largely independent constructs, which may share some descriptive traits, but are otherwise unrelated (e.g., Cain et al., 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009).

Consistent with theories of normal narcissism, the NPI's operational definition of narcissism has been described as "simply recognizing one's own contribution to positive outcomes without overly exaggerating one's accomplishments" (Emmons, 1987, p. 16). However, as noted by Raskin and Terry (1988), such a broad conception of narcissism runs the risk of rendering the term a vague and somewhat meaningless way of describing "all human efforts" (see Ellis, 1927; Hoffman, 2008). It also makes it difficult, if not impossible, to differentiate narcissism from definitions of self-esteem (Pulver, 1970; Westen, 1990), such as "the individual ... is aware of his virtues ... [and has] confident anticipation of success" (Rosenberg, 1965, p. 31).

Moreover, the NPI was not developed to measure historical normative conceptualizations of narcissism. Therefore, such theories are not particularly germane to a discussion of NPI-based research. Instead, the NPI was intended to measure less extreme levels of narcissism "as defined by" a specific clinical typology, the DSM-III criteria for NPD (Raskin & Hall, 1979, p. 590). Thus, it should not be necessary or desirable to segregate NPI-based research from clinical narcissism research, as has been suggested elsewhere (Cain et al., 2008; Pincus & Lukowitsky, 2010; Pincus et al., 2009). Our results indicate that taking a critical approach to the NPI—by recognizing that it contains poorly-conceived items that appear to be confounded with self-esteem—may make it easier to understand how NPI-based research and clinical narcissism research might yet form a "cohesive knowledge base" that converges on a unitary narcissism construct (see Pincus & Lukowitsky, 2010, p. 422).

However, it is also possible that a DSM-based definition of narcissism was too narrow a guide for developing the NPI and/or is too limited as a benchmark measure against which to test the NPI. There is more than a century's worth of theory describing narcissism in ways that reach well beyond the short list of diagnostic criteria included in the versions of the DSM published during the last 30 years (see Morrison, 1986; Ronningstam, 1998; Ronningstam & Gunderson, 1988). However, essential criteria contained in the DSM, such as arrogance, grandiosity, entitlement, exploitativeness, and lack of empathy, are also at the core of the vast majority of other typologies of narcissism. And in the case of all but the most normative theories, narcissists' self-esteem is thought to be either low, or at best, disordered (Westen, 1990), rather than high. Because of this, although using the DSM as a research template may omit some important aspects of narcissism, conclusions reached in narcissism research using the NPI, or any other measure, should still stand up to scrutiny through the lens of the DSM's criteria.

But where does a focus on less extreme levels of characteristics associated with clinical narcissism leave the types of healthy,
normal narcissism recognized by Freud (1931/1950) and Kohut (1966), and referenced by social-personality psychologists (e.g., W. K. Campbell, 2001; Miller & Campbell, 2008)? Do some or all of the NPI’s seemingly poorer items measure this other construct? Without a clear definition of normal narcissism, or strong evidence that it exists independent of other constructs, this question is difficult to answer. However, if research on normal narcissism can be more accurately described as investigations of self-esteem (e.g., Rosenberg, 1965), assertiveness (e.g., Rathus, 1973), or leadership motivations (e.g., Avolio & Locke, 2002; Gough, 1969), etc., it might be preferable to identify and measure those constructs accordingly, leaving the narcissism moniker behind.

Miller et al. (2009), take an opposing view. They suggest that a DSM-based benchmark is inherently limited. Instead, they recommend that the normative characteristics measured by the NPI, such as assertiveness and extraversion, should be incorporated more broadly into the operational definition of narcissism. Further, they point out that such characteristics are already considered by clinicians to be important aspects of narcissism (e.g., Lynam & Widiger, 2001; Samuel & Widiger, 2004). However, as noted earlier, such characteristics exhibit low specificity—they most likely do not differentiate well between narcissists and non-narcissists. And thus, although narcissism may be associated with the presence of these characteristics, they may be just as prevalent in non-narcissists. Because of this, they are not good candidates for inclusion in a narcissism scale that aims for discriminant validity in its predictions. The effects of measuring narcissism based on the NPI’s overinclusive definition of the trait were borne out in the descriptive data of Miller et al. (2009), which indicated that their normal undergraduate sample had significantly higher NPI scores than did their clinical outpatient sample ($t(95) = 2.59, p < .01$). In contrast, as would be expected, their clinical sample had significantly higher narcissism scores on clinical (i.e., SCID-II) interviews (First et al., 1997) than did their undergraduate sample ($t(95) = 2.89, p < .01$). The fact that patients who are higher in clinically-defined narcissism can have lower NPI scores than do normal undergraduates further underscores the likelihood that some of the characteristics measured by the NPI do not help determine whether an individual is actually narcissistic.

8.2. Broader implications?

Reevaluating the NPI—assessing whether the scale’s least narcissistic items measure self-esteem better than they measure narcissism—may shed new, albeit speculative, light on a number of current debates about narcissism and health at a societal level. For example, there is a dispute about whether a steady increase in aggregated NPI scores over the past few decades signals a problematic rise in egotism and self-centeredness among college students (see Twenge, Konrath, Foster, Campbell, & Bushman, 2008a, 2008b), or whether NPI subscale data indicate a less worrisome increase in only the least pathological aspects of normal narcissism (see Trzesniewski, Donnellan, & Robins, 2008). However, if reassessed from the standpoint of the current research, increasing scores on certain NPI subscales may not reflect an increase in narcissism at all. They may instead be the result principally of increases in self-esteem and other similar characteristics. If this were the case, the research may reflect a positive, rather than negative societal change.

Similarly, Foster, Campbell, and Twenge (2003) appeared to support Lasch’s (1979) assertion that western societies are more narcissistic than eastern societies by finding that scores on the aggregated NPI and some NPI subscales were higher in the west than in the east. However, they also reported that there were no cultural differences in scores on the most clearly narcissistic NPI subscales (in this case, the Entitlement and Exploitativeness sub-scales from the 40-item NPI developed by Raskin & Terry, 1988). From our perspective, these data may point to a counterintuitive result—core characteristics of narcissism may be just as prevalent in the east as in the west. Because of confusion about what the NPI measures, we suggest caution in making broad assertions regarding NPI-based findings, including in the two debates referenced here.

8.3. Limitations

Although we focus on the NPI’s overlap with self-esteem, self-esteem is not the only trait with which the NPI is confounded. Research indicates that the NPI may also fail to discriminate between narcissism and other normative traits including dominance (Brown & Zeigler-Hill, 2004) and extraversion (Emmons, 1984), and that L/A in particular may be confounded with assertiveness (P. J. Watson, McKinney, et al., 1988). However, our findings suggest that when the universe is constrained to two self-regard constructs—narcissism and self-esteem—the problematic NPI items that we have identified are more closely aligned theoretically and empirically with self-esteem than with narcissism.

We also recognize that critiquing the empirical overlap of narcissism and self-esteem may oversimplify the relationship between the two constructs. Most clinical theory suggests that narcissism and self-esteem are not independent. Instead, there is thought to be a complex, dynamic relationship between them, with narcissism functioning as a regulatory system that compensates for temporary and chronic deficits in self-esteem (e.g., Akhtar & Thomson, 1982; Kernberg, 1975; Kohut, 1971, 1977; Solorow, 1975; Westen, 1990). This perspective has gained support in the research literature as well (e.g., Morf & Rhodewalt, 2001; Raskin et al., 1991b; Robins, Tracy, & Shaver, 2001; Tracy & Robins, 2003; Tracy et al., 2009). We do not take issue with this view. However, we contend that in order to understand the relationship between two constructs, especially closely related ones, it is critical to measure them in ways that clearly delineate both their core aspects and their boundaries (Clark & Watson, 1995; see also Tracy et al., 2009). Therefore, this article emphasizes the importance of recognizing the differences between narcissism and self-esteem.

It is also important to note that others may take a more nuanced approach to defining high self-esteem, particularly as it relates to narcissism. Some researchers have described the high self-esteem that is linked robustly with the NPI as qualitatively different from non-narcissists’ high self-esteem (Tracy et al., 2009). For instance, narcissists may be described as having accompanying low implicit self-esteem (Bosson, Brown, Zeigler-Hill, & Swann, 2003; Jordan, Spencer, Zanna, Hoshino-Brown, & Correll, 2003; Sakellaropoulos & Baldwin, 2007; Zeigler-Hill, 2006), unstable high self-esteem (Kernis, 2001, 2003), self-esteem that is highly contingent on positive social feedback (Morf & Rhodewalt, 2001), or egocentric grandiosity that only mimics high self-esteem (Westen, 1990). At other times, the relationship between narcissism and self-esteem is described as we have described it here—as a flaw in the NPI with self-esteem confounding the results of NPI-based research (Brown & Bosson, 2001; Paulhus, 1998, 2001; Paulhus et al., 2004). However, most NPI-based research ignores the overlap and potential confound of narcissism and self-esteem, or treats it as a meaningful mediating relationship. Other influential work even balks at making a clear distinction between the two constructs (e.g., Baumeister et al., 2003). This work, in particular, has entered mainstream thought as the basis for confusing news media reports (e.g., Goode, 2002; Sullivan, 2002).

Finally, we recognize that the conclusions reached in this article were based on self-report measures and correlational data. The use of the self-report measure of narcissism (the NPI) was inevitable, given that it is central to the questions being addressed by this
research. However, future research of this type would benefit in particular from the use of both short- and long-term behavioral measures of psychological health and distress.

8.4. Where do we go from here?

Given the problems with the NPI, how might we obtain a more interpretable narcissism measure? There may be a number of ways to retain the NPI by changing how it is used or overhauling it. However, although dividing the NPI based on expert ratings, IRT analyses, and/or factor analyzed subscales helped provide a clear illustration of the scale’s problems in the current research, it may not be the best strategy for its future use. The trouble with this approach is that the resulting NPI-N subscales generally had unacceptably low internal consistency (see also del Rosario & White, 2005; Emmons, 1987; Kansi, 2003; Raskin & Terry, 1988). For instance, the alphas for the NPI-N subscales in Studies 1 and 2 range from .62 to .75. Given the number of items in each NPI-N subscale (18–28), such low reliabilities suggest that these subscales may not measure coherent constructs. Updated factor analyses of the NPI have improved the subscales’ psychometrics (Corry et al., 2008; Kubarych et al., 2004). However, these new subscales are not likely to correct the broader problems created by the inclusion of poor items in the scale.

Better options include controlling for self-esteem (as suggested by Brown & Bosson, 2001; Paulhus, 1998, 2001; Paulhus et al., 2004) or using a revised version of the NPI. The results of Study 2 indicate that controlling for self-esteem yields results that align with those produced by the NPI-N subscales. However, it would clearly be preferable to use a narcissism scale that works properly on its own without requiring statistical manipulation. Alternatively, Ames, Rose, and Anderson (2006) developed a 16-item version of the NPI. They eliminated items that most likely overlapped with more normative constructs such as assertiveness and self-efficacy. The results of their procedure are reflected in our expert rater with more normative constructs such as assertiveness and self-efficacy. The research and manuscript preparation were

Another alternative, the Pathological Narcissism Inventory (PNI; Pincus et al., 2009), was developed to measure multiple facets of pathological narcissism as they range from subclinical to clinical levels. As a measure that encompasses subclinical levels of pathological narcissistic traits, the PNI actually appears to accomplish what the authors of the NPI originally intended. The PNI measures both grandiose (i.e., entitlement rage, exploitativeness, grandiose fantasy, and self-sacrificing self-enhancement) and vulnerable (i.e., contingent self-esteem, hiding the self, and devaluing) aspects of pathological narcissism. Its grandiose subscales (which are the subscales most relevant to a discussion of the NPI) overlap positively with the NPI. But unlike the NPI, they are related to psychological distress (e.g., shame, identity diffusion, interpersonal distress, suicidality) rather than health. They also generally have negligible or negative correlations with self-esteem. It will be important to directly compare the PNI's Entitlement and Grandiose subscales to the Psychological Entitlement Scale and the Narcissistic Grandiosity Scale respectively to determine whether their differing content may provide important incremental information in defining the full spectrum of core aspects of narcissism. Using these three new narcissism scales together may provide a complementary and comprehensive assessment of the range of narcissistic characteristics and problems in normal samples and could begin to make the NPI obsolete. It may also help clarify whether there are any specific narcissistic antecedents of psychological health.

9. Conclusion

Ultimately, the goal of any scale is “to measure one thing (i.e., the target construct)—and only this thing—as precisely as possible” (Clark & Watson, 1995, p. 315). To the extent that there is no definitive answer to the question of what constitutes narcissism and what does not (Millon, 1998; Westen, 1990), this goal may be difficult to achieve. Most psychologists would agree that narcissists are entitled, exploitative, grandiose, self-absorbed, and so forth. Unfortunately, some inferences made in social–personality research are driven by the NPI’s relationship with characteristics that are at best tangential to these domains, such as self-esteem, assertiveness, and leadership motivations. This is evidenced by the fact that the item “I am assertive” was considered narcissistic by only one expert rater in Study 1.

Our concern is that the NPI, and the construct it was developed to measure (i.e., narcissism), have developed a tail-wags-the-dog relationship in the social–personality literature. The NPI is currently the most frequent means for measuring narcissism in psychology research, and an aggregated score is the “most common usage” of the NPI (Miller & Campbell, 2008, p. 456). Because of this, it often appears that narcissism is no longer understood as a latent personality trait estimated by scores on the NPI. Instead, narcissism has become increasingly redefined as “that which the aggregated NPI measures.” Extensive use of the NPI has led to a situation in which social–personality and clinical narcissism research are increasingly difficult to integrate. It also has created the potential for reaching questionable conclusions about narcissism, particularly about its relationship to psychological health.

Acknowledgments

The authors are grateful to Matt Montoya, Diane Purvin, Jennifer Green, and Elsa Ronningstam for their instructive comments. We also thank Yulia Steshenko, Eleftheria Kyriakopoulos, Andrea Hsu Roe, and Lindsey Yourman for their help collecting and preparing data, and 19 anonymous clinical researchers for their time and expertise. The research and manuscript preparation were
supported by the Allport, Ditmars, and Biddle Memorial Research Funds from the Harvard University Department of Psychology, and by a research fellowship from the Center for Public Leadership at the Harvard Kennedy School.

References


