JOURNAL OF RESEARCH IN PERSONALITY **32,** 131–155 (1998)
ARTICLE NO. RP972211

# Idiographic Self-Evaluation and Bias

Rashmi R. Sinha and Joachim Krueger

*Brown University*

Idiographic self-evaluation was conceptualized as the association between a rater's perceptions of trait descriptiveness and that rater's perceptions of trait desirability. In Study 1, self-evaluation was assessed using the 44 items of the Big Five Inventory (John & Donahue, 1994). A subset of 23 items, like the full inventory, showed that most raters' self-evaluations were positive and stable. Study 2 showed that self-evaluation predicted self-esteem and (inversely) depression but not impression management or self-deception. Narcissism was negatively related to self-evaluation when self-esteem was controlled. An idiographic index of evaluative bias (self-enhancement or self-diminishment) was derived from the self-evaluation index by partialing out the group averages of descriptiveness and desirability ratings. In Study 3, this index of bias was compared with two difference-score indices: (1) the degree to which people rated themselves more or less favorably than they rated others and (2) the degree to which they rated themselves more or less favorably than they were rated by others. The idiographic index was independent of the two difference-score measures and showed greater self-enhancement. 

Jennifer considers herself creative and friendly; Joan thinks of herself as shy and nervous. Both self-concepts have descriptive and evaluative aspects. The descriptive aspect captures the content of the self-concept in terms of the person's perceptions of her typical behaviors. Jennifer is likely, or believes herself likely, to generate new ideas and to make friends easily. Joan may need time to warm up to others and her behaviors may express unease. Such differences in self-reported behaviors and traits are central to the assessment of individual differences in personality. In contrast, the evaluative aspect captures the overall favorability of the self-concept. Jennifer has a more positive view of herself than Joan does. Such differences in global self-evaluation are central to the assessment of individual differences in well-being. As shown in the example, the two perspectives on assessment emphasize

We are grateful to Brian Hayden, Liz Tighe, and two anonymous reviewers for providing many helpful comments on a draft version of this manuscript.

Address correspondence and reprint requests to Joachim Krueger, Department of Psychology, Box 1853, Brown University, Providence, RI 02912. E-mail: Joachim_Krueger@Brown.Edu.



0092-6566/98 $25.00

different aspects of the self-concept, but these aspects are not independent because they are inferred from the same set of adjectives.

Personality and clinical assessment have responded differently to the linkages between self-description and self-evaluation. In personality assessment, some researchers strive to purify self-descriptions by controlling differences in self-evaluation. Raters who endorse positive rather than negative statements are thought to respond, at least in part, to the social desirability of the inventory items rather than their content. Individual differences in socially desirable responding may obscure genuine differences in self-description (Paulhus, 1991). According to this view, Jennifer's self-description is suspect because of its positivity. Measurement instruments have been developed to detect (and discard) the offenders (Crowne & Marlowe, 1964; Meehl & Hathaway, 1946). In clinical assessment, some researchers strive to purify self-evaluations by minimizing differences in self-description. Positive global self-evaluations are seen as a key antecedent of mental health and psychological adjustment. Raters who endorse positive rather than negative statements are thought to be high in self-esteem (Rogers, 1961; Rosenberg, 1979). According to this view, Jennifer's self-description approaches the ideal; Joan may require therapeutic intervention.

There is yet a third research perspective which posits automatic and illusory biases in self-perception (Taylor & Brown, 1988). Most work conducted from this perspective has been done by social psychologists who were less concerned with the level of self-evaluation than with the discrepancies between self-evaluation and normative criteria for what that evaluation should be. One criterion is the person's evaluation of the ''average other'' (e.g., Brown, 1986; Heckhausen & Krueger, 1993). In this ''common-rater paradigm'' (CRP), the critical comparison is within the person. Another criterion is how positively the person is evaluated by others (Campbell & Fehr, 1990; John & Robins, 1994). In this ''common-target paradigm'' (CTP), the critical comparison is between persons. Although these two paradigms subtract different criteria from self-evaluations, they yield similar results. When difference scores indicate bias, positive self-evaluations imply self-enhancement and negative self-evaluations imply self-diminishment. The more positively raters describe themselves, the more likely they are to describe others less positively (common rater) and be described by others less positively (common target). Despite this similarity in outcome, the interpretation of the bias varies with the paradigm. Researchers using the CRP have concluded that self-enhancement supports mental health (e.g., Taylor & Gollwitzer, 1995), whereas researchers using the CTP have concluded that self-enhancement undermines mental health (e.g., Colvin, Block, & Funder, 1995). This difference in interpretation mirrors the different perspectives used in clinical and personality assessment.

*Normative Desirability*

Despite their differences in interpretation, all three research perspectives share the assumption that the desirability of trait-descriptive adjectives can be sufficiently represented by their consensual or ''social'' desirability. A socially desirable trait is a trait that raters, as a group, judge to be desirable. Individual raters may not be perfectly reliable judges of desirability, but as a group they are highly reliable (and valid by definition). Inasmuch as individual differences in desirability ratings for a given trait are random measurement error, aggregation across raters improves the precision of measurement, and trait terms can be ranked according to their scale scores of social desirability (Anderson, 1968; Hampson, Goldberg, & John, 1987).

Given these assumptions, self-evaluation can be indexed by the correlation between ratings of trait descriptiveness and the average (social) desirability scores. In the introductory example, it was assumed that Jennifer and Joan agreed on the degree to which being creative is more desirable than being nervous, and thus Jennifer's self-evaluation was more positive than Joan's self-evaluation. To examine self-enhancement, each woman could be asked to rate the other one, too. The CRP would indicate self-enhancement bias if Jennifer rated the socially desirable traits as being more descriptive of herself than of Joan. The CTP would indicate bias if Joan rated the socially desirable traits as being less descriptive of Jennifer than Jennifer herself did.[1]

The assumption of social desirability is attractive because mean interrater correlations rarely fall below .6 (Krueger, 1996a, 1998a; Rothbart & Park, 1986). Individuals' ratings are highly correlated with group averages (Edwards, 1965). Because trait desirability ratings are in largely part stimulus-driven, Joan can be expected to agree with Jennifer that it is more desirable to be creative and friendly than it is to be shy and nervous.

*Idiographic Desirability*

There are, however, limitations to the raters' consensus on trait desirability. Social perception is, in part, an active process reflecting the state of the perceiver (Alicke, 1993). Most evaluative judgments not only depend on objective worth but also on subjective utility. A dollar looms larger to the poor than to the rich. Differences in evaluation also reflect differences in attitude. People who hold a certain belief probably consider this belief to be more desirable than do those who do not hold it. Supporters of free trade believe that free trade is good—and that is why they support it. Supporters of protectionism believe that protectionism is good, and so forth. To attitude

[1] The CRP confounds actual differences in personality with biased perceptions. The CTP presents a less ambiguous index of bias by holding the target of description constant.

researchers, the differences between raters are more interesting than the differences in average evaluation between attitude objects. To say that free trade is more popular or socially desirable than protectionism may simply mean that there are more free traders than protectionists.

Abelson (1986) suggested that attitudes are mental possessions and are evaluated as such. Possession itself can enhance evaluation (Bar-Hillel & Neter, 1996). The same reasoning may apply to personality traits. One source of enhancement is a boost in the perceived importance of a trait once it is possessed (Lewicki, 1983). For example, students who received good rather than poor grades in a computer science class felt that computer skills were an important basis for judging others (Hill, Smith, & Lewicki, 1989). This example shows that to understand self-evaluation, it is not sufficient to consider differences in evaluation between characteristics (i.e., social desirability) but one must also consider differences between raters' perceptions of the same characteristic.

A direct reflection of individual differences in trait evaluation is differences in ratings of desirability. Horney (1950) suggested that

> . . . what is a shining asset to one person is a disgraceful liability to another. One person is proud of being rude to people; another is ashamed of anything that could be construed of as rudeness and is proud of his sensitiveness. (p. 93)

Horney's hypothesis was too strong because it implied that the normative social desirability of traits has little meaning. Most people agree, however, that it is more desirable to be creative or friendly than it is to be shy or nervous. Nevertheless, those who (feel they) possess a certain trait may view it as more desirable than those who do not possess it. A recent investigation supported this idea (Krueger, 1998a). Although raters showed substantial agreement in discriminating between desirable and undesirable traits, raters who considered a given trait to be self-descriptive rated it as more desirable than raters who did not consider it to be self-descriptive.

An idiographic measure of self-evaluation needs to take into account that ratings of trait desirability are not only affected by the trait's socially normative desirability but also by whether the rater considers the trait to be self-descriptive. The correlation between a person's ratings of trait descriptiveness and that person's ratings of trait desirability is a straightforward index of self-evaluation (see also Pelham, 1993, or Pelham & Swann, 1989, for a similar measure involving importance ratings). But on which traits should the index be based? One approach is to decontextualize measurement by selecting items from a broad spectrum of individual differences. Alternatively, items can be selected to represent specific social settings. Studies 1 and 2 apply the first approach, and Study 3 applies the second one.

The goal of Study 1 was to measure self-evaluation idiographically by using a broad item pool. Subsequently, we asked how well individual differ-

ences in self-evaluation could be reproduced by using a smaller sample of items. Study 2 examined how well self-evaluation predicted self-esteem, narcissism, and other relevant constructs. Study 3 used an item pool appropriate for small interacting groups. Its main goal, however, was to revisit the issue of bias. We examined individual differences in bias by assessing self-evaluation while controlling the group averages of the descriptiveness ratings and the group averages of the desirability ratings (see Krueger, 1998a, for a description of the properties of this measure). The resulting partial correlations capture the degree to which raters describe themselves more positively (self-enhancement) or more negatively (self-diminishment) than the social norms suggest (see also Krueger, 1996a). Study 3 examined the relationship between the idiographic index of bias and difference scores derived from the CRP and the CTP.

## STUDY 1: MEASUREMENT

The purpose of Study 1 was to develop a procedure for the reliable measurement of individual differences in self-evaluation. The 44 trait descriptions of the ''Big Five Inventory'' (BFI) formed the initial item pool (John & Donahue, 1994). Participants rated each trait both as to how well it described them and how desirable it was. A self-evaluation index was computed by correlating the participant's descriptiveness ratings with his or her desirability ratings across traits. The average positivity, its variability, and reliability were examined. Next, using the cross products of descriptiveness and desirability ratings, it was examined whether a subset of items would suffice for reliable measurement. Self-evaluation scores were then computed only across the most reliable items. Bias scores were also computed as partial correlations for both the full inventory and the subset of the most reliable items.

### Method

Forty-two male and 56 female undergraduate students (average age = 18.49 years) participated in exchange for partial credit in an introductory psychology class. They participated in groups of 8 to 10.

Upon entering the laboratory, participants learned that the study dealt with certain aspects of social judgment. They were given a three-page questionnaire with instructions. Each page listed the 44 trait descriptions of the BFI (e.g., ''I see myself as someone who is talkative''). The BFI is a psychometrically reliable and valid instrument for the assessment of five broad personality dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. Whereas the Big Five personality dimensions may not represent an exhaustive map of individual differences (Block, 1995), they are widely regarded as an adequate taxonomy (Goldberg, 1993; John, 1990). Self-evaluation was assessed across trait descriptions varying in content because the objective of the index was to capture self-evaluation independent of descriptive individual differences.

On separate pages, participants made three ratings for each trait description: (a) self-description (1 = does not describe me, 9 = describes me very well), (b) the estimated percentage

of Brown undergraduates who would rate themselves as having that trait, and (c) desirability (1 = undesirable, 9 = desirable). The first and third set of ratings were relevant for the present study. Upon completion of the ratings, participants were debriefed and thanked.

## Results and Discussion

### *Full-Inventory Analyses*

For analysis, within-rater correlations were transformed to *Z* scores. As expected, self-evaluations were positive ($M = .43$, $p < .001$) and variable across raters ($SD = .38$). The shape of the distribution was close to normal ($p > .20$ by Lilliefors test; Neave & Worthington, 1988). Most importantly, these individual differences were systematic. The split-half reliability was assessed by recomputing self-evaluation correlations separately for the first half (items 1 to 22) and the second half of the inventory (items 23 to 44) and correlating the *Z* scores across raters, $r(96) = .72$, $p < .001$. When computed separately for each BFI dimension, all mean scores were positive ($M_{\text{Extraversion}} = .52$, $M_{\text{Agreeableness}} = 1.05$, $M_{\text{Conscientiousness}} = .63$, $M_{\text{Neuroticism}} = .21$, $M_{\text{Openness}} = .76$) and modestly intercorrelated across raters (ranging from .06 to .40, $M = .21$). These findings suggest that differences in self-description across domains of personality do not explain differences in self-evaluation.

The partial correlations measuring bias (self-evaluation with group averages of the descriptiveness and desirability ratings controlled) showed a small but statistically reliable self-enhancement effect ($M = .08$, $p < .001$). Self-evaluations were more positive than could be predicted from the social norms of trait descriptiveness and desirability. The variability of the bias scores ($SD = .11$) was smaller than the variability of the self-evaluation scores, and the two indices were correlated across raters, $r(96) = .76$, $p < .001$. The bias index was normally distributed ($p > .20$), indicating that self-enhancers and self-diminishers should not be construed as distinct categories of people.

### *Item Selection*

The next goal was to develop an index for self-evaluation by eliminating items from the full-inventory measure while preserving its psychometric properties. Because self-evaluation is a correlational index, some of the standard rules of item selection did not apply. Responses to reliable items are highly correlated with the overall scores, which are usually the summed responses to individual items. In the present case, however, neither the sum of the descriptiveness ratings nor the sum of the desirability ratings by itself could represent the pattern of association between the two variables. Therefore, individual ratings of descriptiveness or desirability were not useful for reliability analysis.

*Cross products.* The cross products of the standardized rating variables satisfy the criteria for reliability analysis.[2] When ratings of descriptiveness and desirability are multiplied for each item, one can ask how well individual items predict the correlation between rated descriptiveness and desirability across items. The computational formula of Pearson's product–moment correlation, which is

$$r = \frac{\sum XY}{\sqrt{N \sum X^2 \, N \sum Y^2}},$$

shows that the sum of the cross products covaries perfectly with the entire ratio across cases (e.g., raters). Because each of the 98 raters produced a cross product of trait descriptiveness and trait desirability for each of the 44 items, the data matrix comprised 4312 entries. In the full-inventory analysis, the mean pairwise correlation computed between traits and across raters was low (mean $r = .08$), but the items formed a reliable composite ($\alpha = .79$). Corrected item-total correlations ranged from $r = .002$ to .58 (mean $r = .26$), with a drop after the 23rd of the ranked items. The cross products of the top 23 items correlated better than .23 with the sums of the cross products. These items were retained for the measurement of self-evaluation (see Appendix).

*Analysis of the 23-item set.* The short index correlated highly with the full-inventory measure across raters, $r(96) = .86$, $p < .001$, and its reliability increased relative to the full-inventory measure (mean pairwise $r = .17$; mean item-total $r = .38$; $\alpha = .81$). Figure 1 shows a scatterplot and the best-fitting line for the regression of the full-inventory scores on the short index ($.68X + .2$, $r^2 = .87$). Although the 23 retained items were unevenly distributed across the five trait domains (Extraversion 2, Agreeableness 5, Conscientiousness 6, Neuroticism 0, and Openness 9 items), they appeared to represent the full-inventory measure quite well. The shape of the distribution was approximately normal ($p > .20$). Self-evaluations were somewhat lower ($M = .33$) and more variable ($SD = .53$) than their full-inventory complements.

The distribution of the idiographic bias index was similar to the one obtained with the full inventory ($M = .11$, $SD = .16$, $p < .001$). It did not deviate from normality ($p > .07$), and again, bias was highly correlated with self-evaluation, $r(96) = .76$, $p < .001$.

*Testing alternative selections.* The use of the cross products for reliability analysis and item selection rested on the idea that associations between description and desirability *within items* predict the correlation between the

[2] Oliver John suggested this analysis.

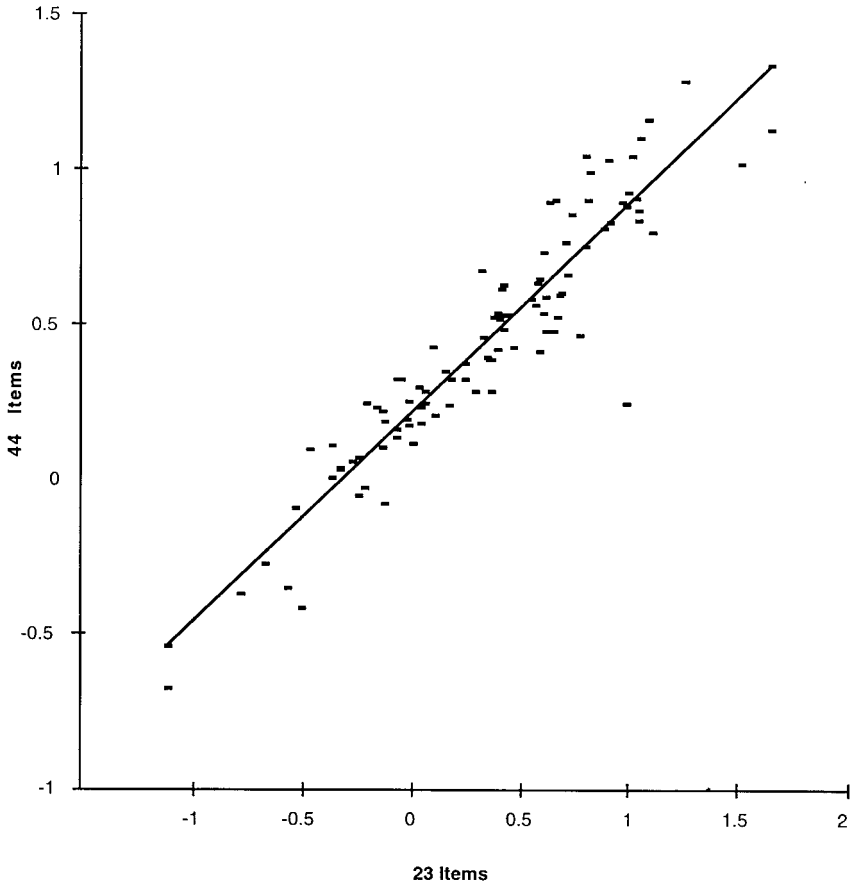


**FIG. 1.** Scores on the 44-item self-evaluation regressed on scores on the 23-item set.

two variables across items. To justify the use of this method, it was necessary to ask whether ratings of descriptiveness or desirability alone could predict self-evaluation. We did this in two ways. First, we asked whether the reliability data (performed on the cross products) could be predicted from either mean descriptiveness or mean desirability ratings alone. Items were rank ordered according to their item-total correlations in the full set of 44 items (as displayed in the Appendix), and this rank order was correlated with the rank orders according to mean descriptiveness ratings ($r(42) = -.04$) and according to mean desirability ratings ($r(42) = -.12$). Second, we asked whether the extremity of the ratings could predict item reliability. Extremity was coded by subtracting all ratings below 5 from 10. Neither the extremity of descriptiveness nor the extremity of social desirability predicted the item-

total correlations ($r(42) = -.27$ and .11, respectively, both $ps > .10$). These two sets of analyses show that a reliable self-evaluation index could not be formed by merely selecting traits with high or low base rates of endorsement or by selecting very positive and very negative traits.

*Stability.* The test–retest reliability of self-evaluation was examined by administering the 23-item index to a new sample of raters twice with an intertest interval of 8 weeks. One hundred sixty-five students (average age = 20.0 years) were recruited in their classes. Fifty-four females and 18 males participated at both times. Mean self-evaluation was stable (Time 1: $M = .58$, $SD = .37$; Time 2: $M = .64$, $SD = .42$), and test–retest reliability was acceptable ($r(70) = .67$, $p < .001$). Bias was also fairly stable (Time 1: $M = .08$, $SD = .18$; Time 2: $M = .07$, $SD = .16$) and reliable, $r(70) = .61$, $p < .001$.

The 23-item index is a brief and easy-to-administer measure of individual differences in self-evaluation. The index was internally consistent and reliable over repeated assessments. A measure of bias was derived from the self-evaluation index by controlling the effects of the social norms (i.e., group averages) of descriptiveness and desirability ratings. Not surprisingly, bias covaried with the self-evaluation. Raters with positive self-evaluations showed greater self-enhancement than raters with negative self-evaluations. On the average, however, bias correlations were lower than self-evaluation correlations, indicating that self-evaluation contains an unbiased (socially valid) component.

## STUDY 2: VALIDATION

Our further examination of the idiographic index of self-evaluation began with an independent assessment of its reliability. Replicating reliability was important because in the data of Study 1, random measurement error may have contaminated the systematic construct-relevant variation. We therefore expected that the reliability of the index would be attenuated in independent testing. The main goal of Study 2, however, was to examine the associations between self-evaluation and conceptually related measures. In particular, the constructs of self-esteem, (lack of) depression, and narcissism involve individual differences in the favorability of the self-concept, and therefore self-evaluation was expected to predict measures of these three constructs. Any measure of favorability may capture a mix of true and false perceptions. A self-concept may be positive because the person possesses desirable characteristics, because the person merely projects an image of desirability, or because of a combination of these factors. We address these possibilities by examining the association between self-evaluation and measures of desirable responding.

In a more exploratory vein, the association between self-evaluation and

the Big Five personality domains was examined. On the one hand, no firm hypotheses were formulated because of the conceptual independence of self-evaluation and the Big Five scores. Self-evaluation captures the association between two variables (description and desirability), whereas scores on the Big Five are aggregated ratings of descriptiveness. Self-evaluation incorporates items from 4 of the 5 domains, and Study 1 had shown similar self-evaluation scores when items from different trait domains were used. On the other hand, it is reasonable to speculate that positive self-evaluation is associated with the presence of positive affect and the lack of negative affect. Positive affect and negative affect have been identified with extraversion and neuroticism, respectively (Watson, Clark, & Tellegen, 1988). Therefore, self-evaluation may be positively correlated with the former and negatively correlated with the latter.

## Method

### *Sample*

Thirty-nine male and 50 female undergraduate students (average age = 18.0 years) participated in exchange for partial credit in an introductory psychology course. They were run individually or in groups of up to 9.

### *Procedure and Materials*

Participants were told that the study addressed certain aspects of social judgment. They were seated in private cubicles equipped with Macintosh IIci computers, which controlled the presentation of instructions and questionnaires and data collection. The items of each of five questionnaires were presented one at a time on the computer screen along with the available response options. Using a mouse, participants indicated their response by pointing the cursor at the chosen alternatives and clicking the button. The order of the questionnaires was randomized across raters but the order of items within questionnaires was constant.

*Self-evaluation and the BFI.* For each of the 23 traits comprising the self-evaluation index, participants rated descriptiveness on a scale from 1 (does not describe me) to 9 (describes me very well) and desirability on a scale from 1 (very undesirable) to 9 (very desirable). They also made descriptiveness ratings for the remaining 21 BFI items so that a complete profile across the Big Five trait domains could be created for each rater.

*Rosenberg Self-Esteem Scale (RSE).* The 10-item RSE (Rosenberg, 1965) measures global feelings of self worth (e.g., ''On the whole, I am satisfied with myself''). Each statement was accompanied by a rating scale ranging from 1 (strongly disagree) to 4 (strongly agree). Because both the RSE and the self-evaluation index tap the positivity of the self-concept, a positive correlation between the two was expected. Note, however, that in contrast to the RSE, self-evaluation focuses on responses to specific trait-related personality descriptors. Self-esteem is thus more likely to reflect generalized affect toward the self, whereas self-evaluation is more likely to reflect concrete recollections of behavioral episodes. This difference in focus may limit the size of their intercorrelation.

*Beck Depression Inventory (BDI).* A negative self-concept is central to the depressive syndrome. The BDI is the most widely used scale for its assessment (Beck, 1967). It has 21 items consisting of 4 or 5 alternative statements each (e.g., ''I feel blue or sad'') varying in depressive content. For each item, raters select the statement that most accurately describes their mood. The expected negative correlation between the BDI and self-evaluation may, in part, reflect

the variance shared by the BDI and the RSE. Compared with self-evaluation and the RSE, however, the BDI may be more sensitive to temporary mood states. For this reason and because the BDI assesses various aspects of depression unrelated to self-evaluation (e.g., pessimism), the correlation between the BDI and self-evaluation was expected to be negative and intermediate in strength.

*Narcissism Personality Inventory (NPI).* The NPI is a 40-item personality inventory, designed to measure narcissism in nonclinical populations. For each pair of statements, raters choose the more self-descriptive one (e.g., ''A: I am assertive. B: I wish I were more assertive.'' Raskin, Novacek, & Hogan, 1991). High scorers tend to be energetic, extraverted, experience-seeking, self-confident individuals whose self-esteem is often high but fragile. They also tend to have grandiose conceptions of themselves and to overestimate their accomplishments (John & Robins, 1994). ''Typically [they are] seen by others as being egotistical and conceited'' (Raskin & Novacek, 1989, p. 67). In the present study, positive correlations between the NPI and self-evaluation were expected.

*Balanced Inventory of Desirable Responding (BIDR).* The BIDR measures the tendency to respond in a socially desirable manner (Paulhus, 1984, 1991). The subscale of self-deceptive enhancement (SDE) refers to an unconscious tendency to overattribute positive attributes to the self (e.g., ''I never regret my decisions''), whereas the subscale of impression management (IM) refers to a conscious tendency to overreport positive attributes and to underreport undesirable ones (e.g., ''I never read sexy books or magazines''). Ratings were made on scales ranging from 1 (not true) to 7 (very true).

SDE is positively related to self-esteem (Paulhus & Reid, 1991; Winters & Neale, 1985) and other measures of mental health (Linden, Paulhus, & Dobson, 1986; Sackeim & Gur, 1979). Because self-evaluation is expected to be correlated with the RSE, it may also predict scores on the SDE. However, because self-evaluation is thought to reflect, at least in part, the actual positivity of the self, the correlation with SDE may be fairly low. IM correlates with lie scales (e.g., the MMPI Lie Scale), need-for-approval scales (e.g., the Marlowe-Crowne Social Desirability Scale; Paulhus, 1984), and measures of agreeableness (Paulhus, 1991). It does not correlate with self-esteem (Paulhus, 1991). Because self-evaluation mainly reflects private rather than public self-evaluation, it was expected that its correlation with IM would be low.

## Results and Discussion

### *Reliability*

The distribution of self-evaluation scores was similar to the one in Study 1. The average was somewhat higher ($M = .47$, $SD = .32$), but the scores were again distributed normally ($p > .20$). As expected, the scale was internally consistent but the relevant indicators were slightly attenuated relative to Study 1 (mean pairwise $r = .10$; mean item-total $r = .29$; $\alpha = .73$). Two of the three conceptually related scales had slightly higher reliabilities than the self-evaluation index (RSE: mean pairwise $r = .47$; mean item-total $r = .62$, $\alpha = .87$; BDI: mean pairwise $r = .27$; mean item-total $r = .50$, $\alpha = .88$); two scales were similar in reliability (NPI: mean pairwise $r = .11$; mean item-total $r = .30$, $\alpha = .82$; IM: mean pairwise $r = .13$; mean item-total $r = .31$, $\alpha = .75$); and one scale was less reliable (SDE: mean pairwise $r = .09$; mean item-total $r = .24$, $\alpha = .65$). Taken together, these data replicate the reliability of the self-evaluation index and they demonstrate that its reliability is as good as that of other, established measures in this domain.

TABLE 1A
Correlations among Self-Evaluation and Related Scales

| | Self-evaluation | RSE | BDI | NPI | SDE |
|---|---|---|---|---|---|
| Self-esteem (RSE) | .43 | | | | |
| Depression (BDI) | −.49 | −.71 | | | |
| Narcissism (NPI) | −.05 | .39 | −.10 | | |
| Self-deception (SDE) | .18 | .41 | −.24 | .39 | |
| Impression-management (IM) | .23 | .20 | −.24 | −.13 | .27 |

TABLE 1B
Correlations between Self-Evaluation and the BFI Scales

| BFI trait domain | Self-evaluation |
|---|---|
| Extraversion | .21 |
| Agreeableness | .25 |
| Conscientiousness | .41 |
| Neuroticism | −.05 |
| Openness | .28 |

*Note.* With $df = 87$, $p < .05$ for $r > .20$, $p < .01$ for $r > .26$, and $p < .001$ for $r > .34$.

### *Validity*

To examine the convergent and discriminant validity of the self-evaluation index, the scale scores derived from the various measures were intercorrelated across raters. As shown in Table 1A, the pattern of correlations was coherent. As expected, self-evaluation was positively correlated with self-esteem and negatively correlated with depression, but it was uncorrelated with narcissism. Because the interrelationships between self-evaluation, self-esteem, and narcissism were of primary theoretical interest, they were examined further. The positive correlation between self-esteem and narcissism is well-documented (e.g., Robins & John, 1997), and it was replicated here. This correlation was important because it raised questions about the unique, unmediated associations between self-evaluation and the two other individual difference measures. Partial correlations revealed an interesting asymmetry. Self-evaluation was related to self-esteem even when differences in narcissism were controlled (partial $r = .45$, $p < .001$). In contrast, self-evaluation was negatively related to narcissism when self-esteem was controlled (partial $r = -.26$, $p < .05$). These findings are suggestive of the dual nature of narcissism. Narcissists may evaluate their own specific characteristics positively only to the extent that they also have high global self-esteem. Otherwise, they may succumb to self-doubts and self-denigration.[3]

[3] We thank an anonymous reviewer for suggesting this analysis and interpretation.

The two facets of socially desirable responding (assessed with the BIDR) predicted only a small part of the variance in the self-evaluation. It is noteworthy that self-evaluation was even less closely associated with self-deception than were self-esteem or narcissism. In other words, the self-evaluation index was the least contaminated by unconscious bias. The remaining intercorrelations among the predictor variables replicated previous findings. Self-esteem was related to all other predictors, as was depression (except for the lack of a correlation with narcissism).

The data in Table 1B show that with the exception of Neuroticism, BFI scores predicted self-evaluation. There was no support for the idea that Extraversion (positive affect) or Neuroticism (negative affect) would best predict self-evaluation. As a composite, however, the BFI scores predicted self-evaluation quite well (multiple $r = .60$). It was therefore necessary to ask whether self-evaluation improved the predictability of criterion measures when differences in the five personality scores are controlled. The five-way partial correlation between self-evaluation and self-esteem was significant ($r(87) = .34$, $p < .001$), and only slightly smaller than zero-order counterpart ($r = .43$). This finding supported the status of self-evaluation as a unique facet of the self-concept.

## STUDY 3: THREE WAYS TO SELF-ENHANCE

Whereas Study 2 focused on the validity of the self-evaluation index, Study 3 focuses on the validity of its derivative, the bias index. It was especially important to determine the relationship between the idiographic bias index and the traditional indices which are based on difference scores. Recall that in both the common-rater paradigm (CRP) and the common-target paradigm (CTP), trait desirability ratings are averaged across raters before bias is assessed. In the CRP, self-enhancement means that raters judge socially desirable traits as more descriptive of themselves than of other people or that they judge socially undesirable traits as less descriptive of themselves than of others. The opposite of these tendencies indicates self-diminishment. In the CTP, ratings of self-descriptiveness are compared with ratings made by others (peers or trained observers).

Instead of computing two separate difference scores to indicate bias, one for socially desirable traits and another one for socially undesirable traits, the present analysis focused on the difference between two correlations. In the CRP, bias is the difference between normative self-evaluation (NSE, the correlation between self-description ratings and average trait desirability ratings) and normative other-evaluation (i.e., the average correlation between ratings of how well the traits describe other individuals and the average trait desirability ratings). A positive difference indicates self-enhancement, whereas a negative difference indicates self-diminishment. In the CTP, bias is the difference between normative self-evaluation and normative peer-

evaluation (i.e., the correlation between the peers' ratings of how well the traits describe the target person and the average trait desirability ratings).

To recapitulate, the bias index of greatest interest in this research was derived from the idiographic self-evaluation measure. It was the correlation between the person's ratings of how well the traits describe him or her and the person's ratings of how desirable the traits are, while the group averages of these sets of ratings were partialed out. One advantage of this index was that it did not depend on ratings of others or ratings by others. Unlike the CRP and the CTP difference scores, the idiographic bias index does not confound self-enhancement with other-diminishment.

Study 3 allowed direct comparisons among the three indices of bias and normative self-evaluation. The outcome of some of these comparisons could be predicted from the computational properties of the indices. First, the idiographic index of bias was designed to be independent of normative self-evaluation. Regardless of whether people endorse mostly socially desirable or undesirable traits as being self-descriptive, the partial correlational index of bias may indicate enhancement or diminishment (Krueger, 1998a). In contrast, the CRP and the CTP indices of bias could be expected to covary with normative self-evaluation. The more positive the NSE is, the more likely it is that other-evaluations or peer-evaluations are less positive. That is, both the CRP and the CTP indices are more likely to yield enhancement biases when the NSE is positive than when it is negative. Second, the CRP and the CTP indices could be expected to be correlated with each other because both were derived from the NSE by subtraction. Because the idiographic index was expected to be independent of the NSE, it was also expected to be independent of the CRP and the CTP indices.

These predictions were tested in a round-robin design where each participant was both a rater and a ratee. The CRP and the CTP were guaranteed to yield the same bias on the mean level because ratings *of* others (other-evaluations) were the same as ratings *by* others (peer-evaluations). Most earlier work showed greater self-enhancement in the CRP than in the CTP. Differences in familiarity and individuation explain this disparity well. In the CRP, most comparisons pit the self against anonymous others, whereas in the CTP, observers typically know the target person well. Supporting this interpretation, studies have shown that self-enhancement is reduced in the CRP when participants compare themselves with other individuals they know well (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Regan, Snyder, & Kassin, 1995). One advantage of the idiographic index was that the amount of observed bias did not depend on the rater's familiarity with others. Moreover, because the idiographic index was sensitive to self-enhancing variations in trait desirability ratings, we expected that this index would yield the strongest self-enhancement effects.

To ensure the relevance of the ratings for the group situation and the behaviors exhibited therein, a new set of items was selected from the BFI.

This procedural modification was expected to increase the reliability and the generalizability of the findings.

## Method

### *Sample*

Forty-four female and 40 male undergraduate students (average age = 18.5 years) participated in groups of 6 in exchange for credit in an introductory psychology course. Only students who did not know any of the other 5 enlisted students qualified for participation. Only data from complete groups with fully completed questionnaires were analyzed. By this criterion, data from three sessions were discarded.

### *Procedures and Materials*

Procedures were modeled after John and Robins's (1994) study. These authors had groups of MBA students discuss hypothetical job applicants until they reached consensus on how to evaluate these applicants. We adapted this task to suit a general student population by presenting topics of broader interest for discussion. Specifically, we simulated the meeting of a student activity committee in a freshman residential unit.

Participants were seated around a table in random order. They received name tags and were encouraged to address each other by first names. The experimenter described the study as an investigation of group problem solving. She handed out written instructions which stated:

> In this experiment you will participate in a group discussion. Imagine that you are living in a freshman unit. Your unit has recently received $500 from the Undergraduate Council of Students, which can be used for any purpose, as decided by the unit members. Further imagine that the others sitting around this table are fellow unit members of a committee whose task it is to decide how to spend the money. As a group you have decided to donate the money to some charity. Each of you, however, will speak to support a different charity. Through discussion, you can arrive at a consensus regarding which charity to support.
>
> You can choose the charity you want to support in this discussion. Your choice, however, is restricted to a list of 10 charities chosen by the experimenter. We will pass the list of charities around the table. Choose the charity you would like to support by placing an ''x'' next to it, and pass the list on to the person on the right. After everyone has made a choice we will give each of you more information about the charity you have chosen.
>
> At the beginning of the group discussion, each of you should introduce yourself, and briefly describe the charity you are supporting. When all six participants have spoken up once, the discussion is open. Remember, your goal is to try to convince the group to support to the charity you have chosen. You can do so by (a) describing the benefits of donating to the particular cause, and (b) critiquing the problems associated with donating to other causes. The goal of the group is to consensually allocate the money. The whole sum of money can be donated to one charity or it can be divided between a maximum of three charities.

All charities included in the list were well known.[4] Participants were given about 40 minutes

[4] Pretesting was done with a group of 6 students. A list of 20 popular American charities was taken from a charity-related site on the World Wide Web. On 10-point scales, group members made two ratings about each charity: ''How much do you know about 'X' '' and ''How much do you like 'X.' '' Ten charities with means within 1 standard deviation on both measures were retained. The charities were Amnesty International, Habitat for Humanity International, National Black Child Development Institute, American Foundation for the Blind,

TABLE 2
The Reliability of Ratings of Trait Descriptiveness and Desirability

| | Descriptiveness | Desirability |
|---|---|---|
| *Within sessions* | | |
| Mean Pairwise *r* | .30 (.41) | .67 (.76) |
| *Across sessions* | | |
| Mean Pairwise *r* | .35 (.95) | .76 (.99) |

*Note.* Averages within sessions encompass all possible correlations (15) between pairs of raters. Averages across sessions encompass all possible 2144 correlations among the 66 raters. Coefficient $\alpha$ is given in parentheses.

for group discussion. Informal observation suggested that they found the task involving and thought provoking. They appeared to feel motivated to contribute to the group discussion. At the end of the allotted time, participants were asked to declare their decision regarding the charity or charities they had decided to support. Then, they individually filled out the questionnaires with the dependent measures.

### *Measures*

Participants rated the performance of each group member (including their own) on 5 desirable (''*original, comes up with new ideas,*'' ''*clever, sharp-witted,*'' ''*enthusiastic,*'' ''*likes to cooperate with others,*'' ''*considerate and accommodating*'') and 4 undesirable characteristics (''*shy, withdrawn or reserved,*'' ''*argumentative,*'' ''*rigid, not compromising,*'' ''*gets nervous easily*''). These items were selected from the BFI to ensure that they covered a range of behaviors relevant for group discussions. Participants made all ratings of descriptiveness on scales from 1 (does not describe performance) to 8 (describes performance very well). Similarly, ratings of desirability could range from 1 (having characteristic was not desirable in the context of the group discussion) to 8 (having characteristic was desirable in the context of the group discussion).

## Results and Discussion

Because intragroup dynamics varied from session to session, the primary units of analysis were the individual interacting groups of 6 (Kenny & Judd, 1986). Preliminary analyses (Table 2) confirmed that consensus was greater regarding the desirability of the traits than regarding their descriptiveness of the self. This was expected because desirability ratings referred to the same traits for all participants, whereas descriptiveness ratings referred to a different target (the self) for each participant.

Main analyses focused on a set of 36 correlations (6 raters $\times$ 6 ratees) for each session. Six of these correlations were self-related, whereas the other

National Committee for the Prevention of Child Abuse, National Committee AIDS Partnership, National Committee for the Homeless, Human Rights Watch, Just Say No International, and National Abortion Rights Action League Foundation.

30 were other-related or peer-related. The term ''other-related'' refers to correlations involving a participant's ratings *of* another person, whereas the term ''peer-related'' refers to correlations involving ratings of the participant as made *by* another person. In a round-robin design, each correlation that is not self-related can be viewed as either other-related or peer-related, depending on whether the perspective of the CRP (the participant as *rater*) or the perspective of the CTP (the participant as *ratee*) is taken.

The first index captured normative self-evaluation (*NSE*) as the correlation between the self-descriptiveness ratings and the averages of the desirability ratings (computed across all participants). The following three indices captured the degree of bias (self-enhancement vs. self-diminishment). First, the idiographic index (*bias*) was the correlation between a rater's descriptiveness ratings and that rater's own desirability ratings after the two sets of group averages were partialed out. Second, the creation of a *CRP* index required the assessment of the average positivity of the rater's descriptions of other group members (average other-evaluation). To do this, each rater's descriptiveness ratings of each other group member were correlated with the average desirability ratings. The average *Z* score was then subtracted from the *Z*-scored NSE. Third, a *CTP* index was created by first assessing how positively each participant was rated by his or her peers (average peer-image). For each participant, descriptiveness ratings by each peer were correlated with average desirability ratings, and the average *Z* score was subtracted from the *Z*-scored NSE. For both the CRP and the CTP, positive differences indicated self-enhancement and negative differences indicated self-diminishment.

As can be seen in Fig. 2, the averages of the NSE, the idiographic bias, and the other/peer-evaluation was positive ($ps < .0001$). There was no difference between average other-evaluation and average peer-evaluation because both indices were based on the same pool of correlations. On the average, the NSE was no more positive than other/peer-evaluations, and thus there were no CRP or CTP enhancement effects ($M = .01$). Supporting the view that the idiographic index is most sensitive to self-enhancement, its average was positive. Yet, its size was smaller than the average NSE, $t(64) = 2.38$, $p < .01$.

Table 3 shows the correlations among all indices across raters. Significance levels are indicated, but the *pattern* of correlations carries most of the information. The first column shows the correlations involving the NSE. The correlation with idiographic bias was relatively low which indicates that bias was not a by-product of the NSE. Regardless of the level of their normative self-evaluation, raters could self-enhance or self-diminish. The correlation between the NSE and other-evaluation may be interpreted as a general consistency bias or as *projection* (Krueger, 1998b). Participants who liked themselves also liked others. The correlation with peer-evaluation indicated a fair

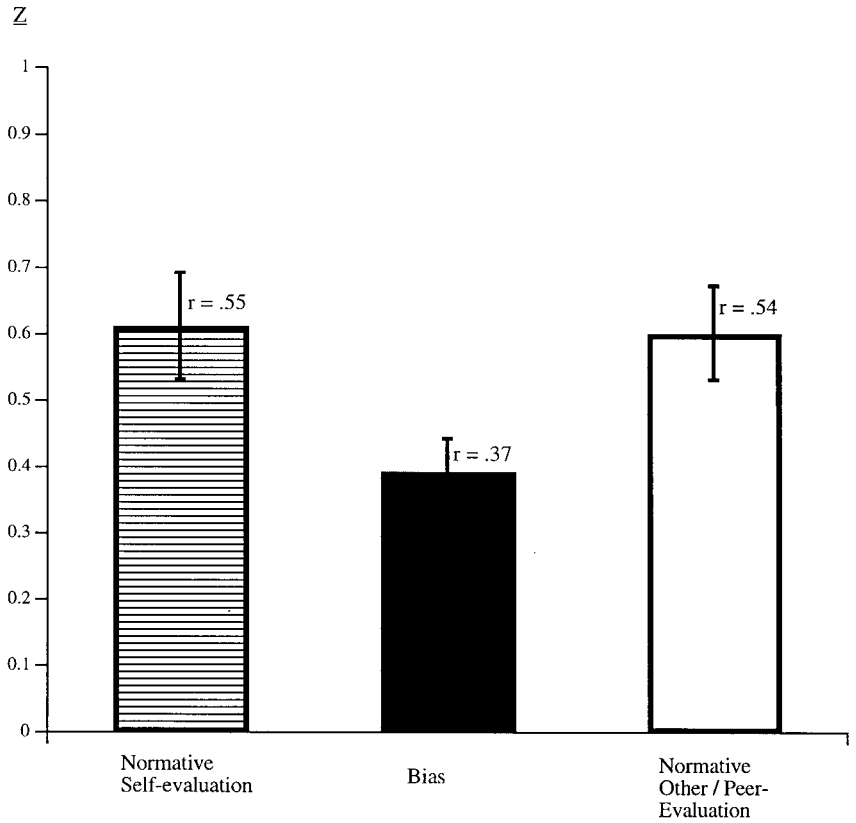


**FIG. 2.** Average indices of positivity and enhancement.

TABLE 3
Correlational Matrix for Various Self Indices (Across 66 Raters)

| | Normative self-evaluation NSE | Bias | Normative other-evaluation | Normative peer-evaluation | CRP |
|---|---|---|---|---|---|
| Bias | .29 | | | | |
| Other-evaluation | .40 | .21 | | | |
| Peer-evaluation | .32 | −.06 | −.12 | | |
| *CRP* | .80 | .17 | −.23 | .42 | |
| *CTP* | .78 | .32 | .47 | −.34 | .52 |

*Note.* With $df = 64$, $p < .05$ for $r > .24$, $p < .01$ for $r > .33$, and $p < .001$ for $r > .40$.

degree of *accuracy*. Participants who liked themselves were also liked by others. The positive correlations between the NSE and the difference score indices of bias were predicted on statistical grounds (see Krueger, 1996b, for a simulation). The second column shows that idiographic bias was only weakly related to other-evaluation, peer-evaluation, and, most importantly, the CRP and the CTP indices. These low correlations supported the expected independence of the idiographic index.

The remaining columns show three noteworthy results. First, other-evaluations and peer-evaluations were unrelated. Participants who tended to rate their fellow group members favorably were rated no more favorably by these members than were participants who tended to rate their fellow group members unfavorably. Second, the CRP and the CTP indices were negatively related to other-evaluations and peer-evaluations, respectively. These negative associations were expected because these indices were difference scores obtained by subtracting other- or peer-evaluations from the NSE. Third, the CRP and the CTP indices were correlated with one another. This correlation was expected because both resulted from subtractions from the NSE.[5] When the NSE was controlled, the CRP and CTP indices were no longer positively correlated ($r = -.28$).

## GENERAL DISCUSSION

The present research yielded three sets of findings. The first set was focused on individual differences in self-evaluation. Idiographic self-evaluation was conceptualized as the association between a person's ratings of trait descriptiveness and that person's ratings of trait desirability. This measure departs from standard practice by assuming that individual differences in desirability ratings for a given trait are partly egocentric and systematic and thus should not be eliminated through aggregation. A brief set of trait terms was sufficient for reliable measurement. To our knowledge, cross products for tests of item reliability have not been used in previous research. This method holds promise for standardizing the measurement of other constructs reflected in within-rater correlations (e.g., projection; Krueger, 1998b).

The second set indicated that, conceptually and empirically, the idiographic self-evaluation measure adds to the assessment of a crucial facet of the self-concept. The pattern of correlations with relevant criterion measures can be summarized as follows: Self-evaluation predicts—but is not the same as—global self-esteem (or its inverse: depression), and its predictive validity

[5] Tests of differences between two dependent correlations are notoriously underpowered. Comparing, for example, the correlation between the two conventional indices (.52) with the correlation between the CTP index and the SE (.32) yielded only a $t(63) = 1.45$, $p < .1$, one-tailed (see Steiger, 1980, for procedural details). Note, however, that the first correlation captures only 27% of the variance, whereas the second captures a mere 10%.

is neither subsumed by the Big Five descriptive personality dimensions nor by desirable response tendencies (BIDR). The interpretation of the correlation between self-evaluation and self-esteem (RSE) is a delicate matter because it requires a resolution between conflicting perspectives. On the one hand, a high correlation was expected because both self-evaluation and self-esteem are conceptually related. In the tradition of William James, the two are identical. According to this view, self-evaluation *is* a measure of self-esteem. Block and Robins (1993) took this view by defining intra-individual correlations between Q-sorted descriptions of the real self and descriptions of the ideal self as a measure of self-esteem. On the other hand, self-evaluation is based on a variety of trait ascriptions, whereas Rosenberg's measure of global self-esteem has few concrete referents (''I feel that I have a number of good qualities''). Positive self-evaluations result, in part, from the tendency to rate a given trait as being more desirable if that trait is also considered to be self-descriptive. In contrast, people with high self-esteem consider themselves to be worthy in a generalized way. Is the medium-sized correlation a compromise that permits the claim that self-evaluation and self-esteem are both similar and different? Such a conclusion cannot stand on one correlation alone; it requires additional support.

If self-evaluation and self-esteem captured the same construct, the lack of a perfect correlation could be attributed to unreliable methods. It would be difficult, however, to attribute differences in the *pattern* of correlations with other variables to unreliability. The differences in correlations with narcissism were especially provocative (Study 2). Self-evaluation was positively related to self-esteem when narcissism was controlled, but it was negatively related to narcissism when self-esteem was controlled. It is to be hoped that future research will examine the replicability of this finding by using multiple indices of each of the three constructs.

The third set of findings was focused on the partial correlations which indicated evaluative bias. Bias was conceptualized as systematically self-enhancing or self-diminishing deviations of a rater's judgments from the social norms for these judgments. To capture such deviations, the group averages of ratings of trait descriptiveness and the group averages of ratings of desirability were partialled out of the idiographic self-evaluation indices. Across raters, bias was closely related to idiographic self-evaluation, but its values were lower. This suggests that a high idiographic self-evaluation can be a combination of a realistically positive view of the self and of an enhancement bias. A third index, introduced in Study 3, was the normative self-evaluation which ignored individual differences in trait desirability ratings. The low correlation between bias and the NSE was expected on conceptual and empirical grounds (Krueger, 1998a).

The construction of the idiographic bias index demanded comparisons with indices derived from the common-rater paradigm (CRP) and the com-

mon-target paradigm (CTP). As predicted, the idiographic index was fairly independent. It is a useful additional tool for the measurement of bias because it taps a previously neglected source of variance in trait desirability ratings. The CRP and CTP indices of bias yielded identical results on the mean level and they were highly correlated across raters. The reason for this pattern was that the CRP and CTP indices, but not the idiographic index, depended on normative self-evaluations. The more positive the NSE was, the more negative were both the CRP and the CTP.

It is intriguing that the equivalence of the CRP and CTP indices has gone unrecognized in the literature. The two paradigms coexist within distinct research communities that have asked different questions. Social psychologists often seek to demonstrate bias on the group level. Claims about unrealistic optimism, illusions of vulnerability, heightened perceptions of own ability and fair-mindedness, to name a few, have all arisen from comparisons between self-ratings and ratings of others. In contrast, personality psychologists make weaker claims about bias and stress the partial accuracy of self-perception. Their conclusions are drawn from studies involving comparisons between self-ratings and ratings by others. The present study suggests that these two paradigms are more closely related than has been realized. Future research may examine the possibility of unifying the study of self-evaluation and its biases while preserving the use of nonredundant measures.

## APPENDIX: THE 44 BFI ITEMS WITH ITEM-TOTAL *r*

| I see myself as someone who . . . | | Item-Total *r* | BFI |
|---|---|---|---|
| 1.* | Is a reliable worker. | .58 | C |
| 2.* | Is sophisticated in art, music, or literature. | .48 | O |
| 3.* | Is sometimes rude to others. | .46 | A |
| 4.* | Prefers work that is routine. | .44 | O |
| 5.* | Tends to be disorganized. | .44 | C |
| 6.* | Likes to reflect, play with ideas. | .43 | O |
| 7.* | Has few artistic interests. | .43 | O |
| 8.* | Tends to be lazy. | .40 | C |
| 9.* | Is curious about many different things. | .39 | O |
| 10.* | Values artistic, aesthetic experiences. | .38 | O |
| 11.* | Does things efficiently. | .37 | C |
| 12.* | Perseveres until the task is finished. | .36 | C |
| 13.* | Is talkative. | .34 | E |
| 14.* | Does a thorough job. | .34 | C |
| 15.* | Is ingenious, a deep thinker. | .33 | O |
| 16.* | Makes plans and follows through with them. | .31 | C |
| 17.* | Has a forgiving nature. | .31 | A |
| 18.* | Is original, comes up with new ideas. | .29 | O |
| 19.* | Is generally trusting. | .29 | A |
| 20.* | Has an active imagination. | .25 | O |
| 21.* | Likes to cooperate with others. | .25 | A |
| 22.* | Is reserved. | .24 | E |
| 23.* | Tends to find fault with others. | .24 | A |
| 24. | Is relaxed, handles stress well. | .21 | N |
| 25. | Can be tense. | .21 | N |
| 26. | Can be moody. | .21 | N |
| 27. | Is easily distracted. | .21 | C |
| 28. | Worries a lot. | .19 | N |
| 29. | Is depressed, blue. | .18 | N |
| 30. | Gets nervous easily. | .18 | N |
| 31. | Starts quarrels with others. | .18 | A |
| 32. | Is full of energy. | .16 | E |
| 33. | Generates a lot of enthusiasm. | .16 | E |
| 34. | Remains calm in tense situations. | .15 | N |
| 35. | Is considerate and kind to almost everyone. | .14 | A |
| 36. | Has an assertive personality. | .13 | E |
| 37. | Is inventive. | .11 | O |
| 38. | Is outgoing, sociable. | .10 | E |
| 39. | Can be somewhat careless. | .07 | C |
| 40. | Can be cold and aloof. | .07 | A |
| 41. | Is emotionally stable, not easily upset. | .06 | N |
| 42. | Is sometimes shy, inhibited. | .06 | E |
| 43. | Is helpful and unselfish with others. | .05 | A |
| 44. | Tends to be quiet. | .002 | E |

## REFERENCES

Abelson, R. P. (1986). Beliefs are like possessions. *Journal for the Theory of Social Behaviour*. **16,** 223–250.

Alicke, M. D. (1993). Egocentric standards of conduct evaluation. *Basic and Applied Social Psychology,* **14,** 171–192.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. (1995). Personal contact, individuation, and the better-than-average-effect. *Journal of Personality and Social Psychology,* **69,** 213–226.

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology,* **9,** 272–279.

Bar-Hillel, M., & Neter, E. (1996). Why are people reluctant to exchange lottery tickets? *Journal of Personality and Social Psychology,* **70,** 17–27.

Beck, A. T. (1967). *Depression: Causes and treatment.* Philadelphia: Univ. of Pennsylvania Press.

Block, J. (1995). A contrarian view of the Five-Factor approach to personality description. *Psychological Bulletin,* **117,** 187–215.

Block, J., & Robins, R. W. (1993). A longitudinal study of consistency and change in self-esteem from early adolescence to early adulthood. *Child Development,* **64,** 909–923.

Brown, J. D. (1986). Evaluation of self and others: Self-enhancement biases in social judgments. *Social Cognition,* **4,** 353–376.

Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology,* **58,** 122–133.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology,* **68,** 1152–1162.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Edwards, A. L. (1965). Measurement of individual differences in ratings of social desirability and in the tendency to give socially desirable responses. *Journal of Experimental Research in Personality,* **1,** 91–98.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist,* **48,** 26–34.

Hampson, S. E., Goldberg, L. R., & John, O. P. (1987). Category-breadth and social-desirability values for 573 personality terms. *European Journal of Personality,* **1,** 241–258.

Heckhausen, J., & Krueger, J. (1993). Developmental expectations for the self and most other people: Age-grading in three functions of social comparisons. *Developmental Psychology,* **29,** 539–548.

Hill, T., Smith, N. D., & Lewicki, P. (1989). The development of self-image bias: A real world demonstration. *Personality and Social Psychology Bulletin,* **15,** 205–211.

Horney, K. (1950). *Neurosis and human growth: The struggle toward self-realization.* New York: Norton.

John, O. P. (1990). The ''Big Five'' factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: The Guilford Press.

John, O. P., & Donahue, E. M. (1994). *The Big Five Inventory: Technical report of the 44-item version.* Institute of Personality Assessment and Research, University of California, Berkeley.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology,* **66,** 206–219.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin,* **99,** 422–431.

Krueger, J. (1996a). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology,* **71,** 536–548.

Krueger, J. (1996b). Probabilistic national stereotypes. *European Journal of Social Psychology,* **26,** 961–980.

Krueger, J. (1998a). Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin,* **24,** 505–516.

Krueger, J. (1998b). On the perception of social consensus. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 163–240). San Diego: Academic Press.

Lewicki, P. (1983). Self-schema and social information processing. *Journal of Personality and Social Psychology,* **47,** 1177–1190.

Linden, W., Paulhus, D. R., & Dobson, K. S. (1986). Effect of response styles on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology,* **54,** 309–313.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology,* **30,** 525–564.

Neave, H. R., & Worthington, P. L. (1988). *Distribution free tests.* Boston, MA: Unwin Hyman Ltd.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology,* **46,** 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes: Measures of social psychological attitudes,* Vol. 1 (pp. 17–59). San Diego, CA: Academic Press.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology,* **60,** 307–317.

Pelham, B. W. (1993). The idiographic nature of human personality: Examples of the idiographic self-concept. *Journal of Personality and Social Psychology,* **64,** 665–677.

Pelham, B. W., & Swann, W. B., Jr. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology,* **57,** 672–680.

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissistic self-esteem management. *Journal of Personality and Social Psychology,* **60,** 911–918.

Raskin, R., & Novacek, J. (1989). An MMPI description of the narcissistic personality. *Journal of Personality Assessment,* **53,** 66–80.

Regan, P. C., Snyder, M., & Kassin, S. M. (1995). Unrealistic optimism: Self-enhancement or person positivity? *Personality and Social Psychology Bulletin,* **21,** 1073–1082.

Robins, R. W., & John, O. P. (1997). Effects of visual perspective and narcissism on self-perception: Is seeing believing? *Psychological Science,* **8,** 37–42.

Rogers, C. R. (1961). *On becoming a person.* Boston, MA: Houghton-Mifflin.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton Univ. Press.

Rosenberg, M. (1979). *Conceiving the self.* New York: Basic Books.

Rothbart, M., & Park, B. (1986). On the confirmability and disconfirmability of trait concepts. *Journal of Personality and Social Psychology,* **50,** 131–142.

Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology,* **47,** 213–215.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin,* **87,** 245–251.

Taylor, S. E., & Brown, J. (1988). Illusion and well-being: A social-psychological perspective on mental health. *Psychological Bulletin,* **103,** 193–210.

Taylor, S. E., & Gollwitzer, P. M. (1995). Effects of mindsets on positive illusions. *Journal of Personality and Social Psychology,* **69,** 213–226.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology,* **54,** 1063–1070.

Winters, K. C., & Neale, J. M. (1985). Mania and low self-esteem. *Journal of Abnormal Psychology,* **94,** 282–290.