Social Desirability and Consensual Validity of Personality Traits

KENN KONSTABEL*, TOIVO AAVIK and JÜRI ALLIK

Department of Psychology, Estonian Centre of Behavioural and Health Sciences, University of Tartu, Estonia

Abstract

The effect of socially desirable responding (SDR) on the consensual validity of personality traits was studied. SDR was operationalized as the sum of items weighted by their respective social desirability values (Social Desirability Index, SDI), which could be computed for both self- and peer-reports. In addition, the Balanced Inventory of Desirable Responding (BIDR) was used as a measure of SDR. It was shown that both self-peer and peer-peer agreement rose significantly for most studied traits when SDI was controlled in both self- and peer-reports. BIDR was a significant suppressor variable in only one of the analyses involving Neuroticism. The SDI detected faking on personality scales somewhat better than the BIDR scales. It is argued that the SDI is a measure of evaluativeness of a person description, and that people agree more on descriptive than on evaluative aspects of a target's personality traits. Copyright © 2006 John Wiley & Sons, Ltd.

Key words: social desirability; personality traits; consensual validity; NEO-PI-R; suppressor variable; moderator analysis

INTRODUCTION

Establishing the validity of the self-reports is one of the principal concerns of research on personality assessment. Objectively measurable validity criteria—for instance frequency counts of behaviours and life outcomes—can hardly ever be thought of as conceptually equivalent to any given personality trait; therefore, an ideal correspondence between the trait scale and the criterion cannot be expected even in the imaginary condition of perfect measurement accuracy. Reports by acquaintances, although based on ordinary social perception just like self-reports, constitute an important validity criterion because their informational basis, as well as the category breadth of the trait descriptors, is at least

*Correspondence to: K. Konstabel, Department of Psychology, University of Tartu, Tiigi 78, Tartu 50410, Estonia.
E-mail: kenn.konstabel@ut.ee

comparable to those of the self-reports. Correlations between the self- and peer-reported personality traits are typically higher than 0.5 (McCrae & Costa, 2003); in some studies, appreciable levels of consensus have been shown even when the raters had minimal information about the target persons (Borkenau & Liebler, 1993; Kenny, Horner, Kashy, & Chu, 1992). The agreement tends to increase as raters have more contact with the persons they are rating (Kenny, 1994); targets' friends are more accurate than strangers (Funder & Colvin, 1997), and spouses are generally more accurate than friends (McCrae & Costa, 2003, p. 42). These studies show that personality ratings are largely consensual, and that the consensus is meaningfully related to the amount of information that the rater has about the target. On the other hand, heightened motivation for socially desirable responding (SDR) has been shown to reduce the consensual validity (Topping & O'Gorman, 1997), and less evaluative scales elicit higher levels of agreement (John & Robins, 1993; Saucier, Ostendorf, & Peabody, 2001). One might expect, then, that controlling for SDR or evaluativeness would enhance self-other agreement. Several studies have shown, however, that traditional social desirability measures fail at this task (e.g. Borkenau & Ostendorf, 1992; McCrae & Costa, 1983): if anything, correcting for social desirability as operationalized by these scales lowered rather than heightened the self-other correlations. There are, however, two reasons to suspect that the previous results may not represent the final truth about the influence of SDR on the consensual validity of trait measures. First, these studies have used social desirability measures that are 'contaminated' with substantive variance (Piedmont, Mcrae, Riemann, & Angleitner, 2000). Second, it has been implicitly assumed that only self-reports may be biased by SDR, and that peer-reports are essentially free of any such biases. The present study aims to revisit the issue by using an index of SDR proposed by Hofstee (2003): the sum of items weighted by their respective social desirability values. This index measures the consistent tendency of a respondent to agree with desirable items and to disagree with undesirable items and can thus be meaningfully computed for both self- and peer-reports.

Operationalization of social desirability

In Wiggins's (1968) discussion of components of variance in personality tests, social desirability is mentioned both as a property of items, and as a response style. Although the present article is concerned only with the response style aspect, it is worth noting that there is a logical correspondence between the two. Assuming a pool of items all keyed in a socially desirable direction, both highly socially desirable items and a highly socially desirable response style can be expected to result in a higher average response and a higher shared component of variance in all items. From this correspondence, it follows that (a) the procedures designed to measure items' social desirability can be adapted to measure socially desirable response style, (b) the manipulations which enhance the motivation for socially desirable responding can be used for measuring the items' and scales' social desirability and (c) individual differences in socially desirable responding should affect scale validity in a similar way as social desirability of the scale items.

   Paulhus (2002) divided the SDR measures into 'minimalist' and 'elaborate': the scales of the first type are comprised of highly evaluative items, typically chosen from a personality scale; the scales of the second type arise from a theoretical elaboration of the construct of social desirability. Examples of the 'minimalist' approach are Edwards' (1957) SD scale composed of the MMPI items with extreme social desirability values, and Wiggins' (1959) SD scale which items best discriminated between 'fake good' and

'accurate description' groups. Based on a similar logic, Schinka, Kinder and Kremer (1997) constructed the validity scales from NEO-PI-R items with extreme mean sores. Crowne and Marlowe's (1960) SDS, an example of an 'elaborate' construct included behaviours that were highly desirable but unlikely to be true. Paulhus' own Balanced Inventory of Socially Desirable Responding (BIDR; Paulhus, 1991) is constructed following the model which divides SDR into impression management (IM) and self-deceptive enhancement (SDE). A later version of BIDR contains an additional scale of self-deceptive denial (SDD) which is, however, strongly correlated with impression management (Paulhus & Reid, 1991). This division, and the evidence that has been found for its support (Paulhus & John, 1998), is beyond the scope of the present paper.

A third, not very often used strategy (not mentioned in Paulhus' review) is to compute social desirability score as a weighted sum of responses to a multidimensional questionnaire. The general idea was probably first proposed by Edwards (1970, pp. 107–108) who defined an SD response as 'a True response to an item with a socially desirable scale value or as a False response to an item with a socially undesirable scale value'. Edwards contended, 'any set of items keyed for SD responses may be regarded as an SD scale' (p. 109), but added a few pages later (pp. 117–118) that for better reliability, items with extreme social desirability values would be preferred. Nevertheless, Comrey and Backer (1975) computed a simple sum of all items of Comrey Personality Inventory (CPI; items keyed in the socially undesirable direction were weighted by -1) and found that this was the best single predictor of faking on CPI. With a different aim, Hofstee and Hendriks (1998) used the sum of Big Five scores (where all scales were keyed in a socially desirable direction) as an index of social desirability, and found a self-other correlation of 0.56. Hofstee (2003, p. 249) proposed that the most obvious measure of social desirability would be the sum of all items weighted proportionally to the items' SD values, where both are expressed as deviations from the neutral midpoint of the scale. To our knowledge, such index has never been actually used. Robins, Tracy, Trzesniewski, Potter, and Gosling (2001) used the sum of items multiplied by their SD values as an index of SDR, but apparently they used untransformed responses and social desirability values in their computations. Consequently, using their approach, giving a neutral response to a neutral item would add more to one's SD score than strongly disagreeing with an extremely undesirable item. Their index thus measures agreeing with desirable items *as opposed to* (rather than 'as well as') disagreeing with the undesirable items, which is counterintuitive.

Another approach to the operationalization of social desirability is experimental or pseudo-experimental manipulation: the participants are provided an incentive to respond in a more socially desirable manner than they would normally do, through a 'fake good' or 'job applicant' instruction. Such manipulations typically produce changes in personality scale scores (e.g. Caldwell-Andrews, Baer, & Berry, 2000; Paulhus, Bruce, & Trapnell, 1995), including social desirability scales (Pauls & Crost, 2004). It has been claimed that the reactivity of SD scales to instructional manipulations is an indicator of their poor validity (Pauls & Crost, 2004), but it seems more natural to assume precisely the contrary: it is a necessary condition of validity for a scale designed to measure SDR that it be sensitive to a manipulation producing a heightened level of SDR.

Social desirability and consensual validity

Surprisingly, the use of SD scales did very little to improve the accuracy of self-report personality measures. Dicken (1963) observed that when social desirability scores

(Edwards's scale) were partialled out the correlation between self-ratings and ratings of knowledgeable others decreased, not increased. Twenty years later McCrae and Costa (1983) confirmed these observations by demonstrating that the correlations between self-spouse ratings on all personality dimensions decreased when the social desirability (Eysenck's Lie and Marlowe-Crowne) was taken into account. Borkenau and Ostendorf (1992) found that the self-other agreement did not increase when controlling for impression management (Lie and Marlow-Crowne scales). Ones, Viswedvaran, and Reis (1996) showed that removing the effect of social desirability from the Big-Five leaves the criterion-related validity of personality practically unchanged. Pauls and Stemmler (2003) found that the correlation between self and other-ratings remained essentially unchanged after adjustment of self-ratings for self-deceptive enhancement and impression management. For Conscientiousness the validity even significantly decreased when self-reports were corrected for scores of impression management. All these results have led to a conclusion that controlling for SD scales such as Marlowe-Crowne and the BIDR is not useful for increasing the validity of personality scales. In many cases, controlling for SD scales even produces significant decreases in self-other agreement.

These results seem to support an interpretation according to which social desirability scales measure substantive personality characteristics rather than a surface style of responding (Block, 1965; McCrae & Costa, 1983; Nicholson & Hogan, 1990; Smith & Ellingson, 2002). Hofstee and Hendriks (1998) emphasize that people are actually behaving in a socially desirable manner (and some people do so more often than others) rather than responding in such way to the questionnaires. Their finding of a considerable self–other agreement on social desirability (0.56) is consistent with that assumption. In a similar vein, Paulhus (1984) has argued that self-deception—a facet of desirable responding according to his model—is an inherent component in some personality characteristics. He did not consider, however, that impression management might be inherent in other personality traits.

On the other hand, there are indications that an SD response style may exist and be related to validity. The shift in scores associated with instructional manipulations would be difficult to explain if socially desirable responding were only substance and no style. Moreover, 'fake good' instructions lower the self-other agreement (Topping & O'Gorman, 1997), and interjudge agreement has been found to be higher in less evaluative scales (John & Robins, 1993; Saucier, Ostendorf, & Peabody, 2001). The failure of a number of studies to show moderator or suppressor effects for SD scales (among others, Borkenau & Ostendorf, 1992; McCrae & Costa, 1983; Ones et al., 1996; Piedmont et al., 2000) may thus show an imperfect validity of these scales rather than the lack of a social desirability effect. Another reason may be that SDR has only been controlled for in the self-reports; no proof has been presented, however, that peer-reports are untouched from social desirability. Indeed, Funder and Colvin (1997, p. 625) have argued that the 'self'-enhancement bias is poorly named because the enhancement effect is also there in comparison between acquaintances' and strangers' ratings. There are several motives that might contribute to the socially desirable reporting about one's acquaintances: for example, the good qualities of one's acquaintance may be perceived as reflecting something about oneself (Tesser, Pilkington, & McIntosh, 1989), or one's friends may be regarded as a part of one's extended self. In addition, socially desirable responses may reflect a benevolent and balanced view of another person, which may contribute to the quality and stability of the relationship (Murray, 1999).

Social desirability: bias or loss of information?

Several studies have investigated the role of SDR as either suppressor or moderator of validity. The choice of moderator or suppressor model, although the reasons for it have rarely been made explicit, has certain theoretical underpinnings. The suppressor model is a natural choice if SDR is viewed as a correctible bias, resulting from, e.g. shifting the 'true response' toward the desirable response (whereas the shift is larger for more evaluative items). On the other hand, when using the moderator model, one supposes that higher levels of SDR imply a loss of information or larger error variance. This may happen, for example, if persons with higher levels of SDR substitute socially desirable response for the 'true response'; in that case, any variation in responses would reflect variations in understanding the social norms rather than variations in personality. The loss of information perspective has been concisely expressed by Comrey and Backer:

The higher the faking predictor score, the less reason there would be to treat the test record as valid. Attempts to construct valid test records by statistical manipulation of invalid responses, however, are not regarded by the authors as very likely to succeed (1975, p. 318).

The present study

The purpose of the present study is to investigate the relationship of SDR with consensual validity, using the measure proposed by Hofstee (2003) as an index of SDR. To emulate the context of testing in applied settings, we include a condition where participants are instructed to respond to the personality tests as they would do if they were applying to a highly desired job. The social desirability index (SDI) computed as the average of all items weighted by their social desirability values (where both are expressed as signed proportions of maximal possible deviations from the scale midpoint) has several advantages over other measures of SDR. First, it has a very simple and intuitive interpretation: the possible values range from -1 to $+1$; positive values indicate a tendency to agree with socially desirable items and to disagree with socially undesirable items, and negative values indicate an opposite tendency. Second, the index can be computed for both peer- and self-reports, and those values can be meaningfully compared. Third, provided a sufficiently heterogeneous item pool, the SDI is expected to be conceptually less dependent on the content of any scale than the social desirability scales.

We will also use another, related method of controlling for the evaluative component of personality ratings that has been proposed by Saucier (1994): partialling item social desirability values from the responses of each individual before computing the scale values. Conceptually, this method should yield similar results as controlling for the SDI as previously described. These two methods are compared to the effect of statistical controlling for a more traditional measure of social desirability, the Balanced Inventory of Socially Desirable Responding (BIDR; Paulhus, 1991). Finally, in addition to suppressor effects, we also examine whether social desirability (either of self- or peer-reports) is a moderator of consensual validity.

METHOD

Participants

The sample included 390 individuals (318 women and 72 men; mean age 22.3 years, SD = 5.2); the respondents were students from University of Tartu, Estonian Agricultural

University, and Mainor Business School. The questionnaires were completed voluntarily; some students studying psychology received an extra credit toward fulfillment of their course requirements. Completed questionnaires were received from approximately 75% of respondents.

All participants were asked to recruit two acquaintances to fill up the test battery about them; in the final analysis, at least one peer-report was available for 376 participants, and two fully complete peer-reports for 289 participants. (Unless otherwise stated, all following analyses are based on average reports of two peers, if two peer-reports are available.) The acquaintances were relatives or close friends of subjects.

Personality measures

The Big-Five dimensions were measured with Estonian version (Kallasmaa, Allik, Realo, & McCrae, 2000) of the NEO-PI-R (Costa & McCrae, 1992) consisting of 240 items forming 30 facet scales and 5 domain scales (Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness). The data were collected in 2002 and 2003.

The Estonian Values Inventory (Aavik & Allik, 2002) was included in the package for another study and will not be discussed further in this paper.

Social desirability scales

Participants also filled up the Estonian version of Balanced Inventory of Desirable Responding (BIDR-6; Paulhus, 1991), which consists of 20 SDE and 20 IM items. The seven-point scale indicates a range from 1 (totally disagree) to 7 (totally agree). Continuous rather than dichotomous scoring scheme was used because the former has been shown to yield higher convergent correlations with other measures of social desirability, and more consistent effects with self-presentation instructions (Stöber, Dette, & Musch, 2002). The inventory was translated into Estonian by Toomas Niit.

Procedure: honest versus applicant instruction

Respondents were assigned to one of the two conditions. In the first condition ('applicant condition'), participants were asked to imagine they are applying for a job they would very much like to have. The instruction, reminded in the header of each page of the questionnaire, stated that:

'You have all skills to perform well in this job, but there are many other candidates whose skills are equal to yours and who are highly motivated to get the job. Please answer the following questionnaires in the way you would do in this situation'.

In the second condition ('honest condition'), the questionnaires were accompanied by a standard instruction to describe themselves honestly and accurately.

There were 226 participants (188 women and 38 men) in the honest condition, 164 participants in the applicant condition (130 women and 34 men). The peer-report instructions were identical in both honest and applicant conditions, containing no reference to the job application situation. Items in the questionnaires used for peer-reports were in third person singular.

Social desirability index

In order to develop a social desirability index for the NEO-PI-R, the questionnaire items were assessed by 88 judges. Judges (24 men and 64 women, mean age 37.6 years, $SD = 12.7$) independently rated the social desirability of each of the 240 NEO-PI-R items. The instruction stated, 'Descriptors of people often contain evaluative information. Some personality characteristics are considered more desirable receiving approval from other people, whereas others are undesirable. If someone agrees strongly with this item - does this present that person in favorable or unfavorable light, or is agreeing with this item neutral as regards to others' approval?' Ratings were made on a 7-point Likert scale ranging from *extremely undesirable* (1) to *extremely desirable* (7), with 4 as neutral. The intraclass correlation reliability of a single judge (ICC (2, 1), Shrout & Fleiss, 1979) was 0.39, and the reliability of the average of 88 judgments was 0.98.

The social desirability index used in the present study was first proposed by Hofstee (2003). The social desirability ratings and participants' responses to each item were first linearly transformed to scale from -1 to $+1$, where 0 denotes a neutral midpoint of the scale (Hofstee & Hendriks, 1998). The transformed responses were then multiplied by the transformed social desirability ratings. In the resulting score, positive values indicate a response in the direction of social desirability; the maximal score, $+1$, would be obtained by completely agreeing with an item with a maximal social desirability value, or completely disagreeing with an item with a minimal social desirability rating. SDI was computed as the average of all weighted items of the NEO PI-R.

Statistical controlling for the SDI

In the results section, we present three different sets of correlations controlled for social desirability. In the terminology used by Cohen (1982), the first two are called *semipartial*, whereas the third one is called *bipartial* correlation. A bipartial correlation differs from the more usual partial correlation in that different sets of variables are partialled out from the 'dependent' and 'independent' variables. Suppose X and Y denote respectively self- and peer-ratings on a personality scale, A and B are the social desirability indices computed respectively on self- and peer-ratings, and '·' denotes partialling (or, equivalently, computing residuals from linear regression). In that case, the correlation between $X \cdot A$ and Y is a semipartial correlation, that between $X \cdot A$ and $Y \cdot B$ is a bipartial correlation, and that between $X \cdot (A,B)$ and $Y \cdot (A,B)$ would be a partial correlation. The latter, however, is not used in the present paper because it would make no sense to partial the SDI of self-ratings from the peer-rated personality scores or vice versa.

RESULTS

Psychometric properties of the social desirability measures

Results on internal reliability of the BIDR scales and their correlations with the NEO PI-R are shown in Table 1. The internal reliabilities of the scales ranged from 0.67 to 0.87, being comparable to the values reported by Paulhus (1991).

The reliability of the SDI was computed as the Cronbach alpha of the 240 NEO PI-R items weighted by their social desirability values. The alphas were 0.95 in the honest condition, 0.97 in the applicant condition and 0.96 in peer-reports. Accordingly,

Table 1.   BIDR and the social desirability index: internal reliabilities and correlations with the NEO PI-R

	Self-deceptive enhancement			Impression management			Social desirability index (SDI)		
	H	A	P	H	A	P	H	A	P
Neuroticism	**−0.64**	**−0.64**	**−0.64**	**−0.27**	**−0.30**	**−0.30**	**−0.75**	**−0.82**	**−0.63**
Extraversion	**0.42**	**0.45**	**0.18**	−0.02	**0.21**	**−0.14**	**0.58**	**0.72**	**0.43**
Openness	0.09	0.12	−0.06	**−0.15**	0.11	−0.07	0.14	**0.39**	**0.26**
Agreeableness	−0.03	−0.03	0.08	**0.58**	**0.43**	**0.60**	**0.46**	**0.41**	**0.58**
Conscientiousness	**0.45**	**0.57**	**0.46**	**0.46**	**0.46**	**0.59**	**0.75**	**0.84**	**0.73**
SDI	**0.59**	**0.56**	**0.50**	**0.45**	**0.47**	**0.56**			
Alpha	0.67	0.72	0.70	0.85	0.84	0.87	0.95	0.97	0.96

Note: The numbers of cases vary, due to missing data, from 210 to 217, in the honest condition, from 153 to 160 in the applicant condition, and from 362 to 365 for the peer-reports.
H = self-reports in honest condition; A = self-reports in applicant condition; P = peer-reports (average of 2 peers).
Significant correlations ($p < 0.05$) are shown in bold type.

approximately 90 items would be needed to achieve a reliability of at least 0.90, and at least 40 items for reliability of at least 0.80. Average intercorrelations between the NEO PI-R scales computed using the weighted items were 0.31 in the honest condition self-reports, 0.44 in the applicants condition, and 0.35 in peer-ratings; all individual correlations were positive. These correlations were higher than the respective correlations (0.17, 0.29 and 0.19) between the original unweighted scales. Neuroticism scores were reversed before analysis so that a majority of correlations would be positive; after the reversal, only three of the inter-domain correlations were negative (the largest one being −0.11 between Openness and Conscientiousness in honest condition self-reports).

The effect of instruction

Table 2 presents the mean scores on personality and social desirability scales for the honest and applicant instruction groups. All of the social desirability indices and all NEO PI-R domain scales except for Openness were significantly different across conditions. As expected, in the applicant condition test takers presented themselves as less neurotic and more conscientious, agreeable, and extraverted compared to the neutral instruction. In addition, from the 30 subscales of the NEO PI-R, 26 demonstrated statistically significant difference; the four subscales that were insensitive to the instructional manipulation were O2:Aesthetics, O3:Feelings; O6:Values and A5:Modesty. The instruction effects on the remaining Openness subscales were different in direction: the mean for O1:Fantasy was lower in the applicant condition, whereas the means for O4:Actions, and O5:Ideas were higher.

Social desirability and item responses

Edwards(1953, p. 92) observed that traits that are judged desirable are those which are fairly widespread or common among people. Indeed, Figure 1 upper panels show that there is a considerable linear trend between social desirability of items and their means in both

Table 2. Personality and social desirability scales: means, standard deviations, and differences between the honest and applicant conditions

	Honest		Applicants		Difference	
	Mean	SD	Mean	SD	t	r_{PB}
Neuroticism	92.7	25.6	68.2	26.4	−9.1	−0.42***
Extraversion	118.7	27.4	132.7	25.7	5.1	0.25***
Openness to Experience	122.1	22.2	124.5	20.2	1.1	0.06
Agreeableness	110.1	20.2	118.7	18.4	4.3	0.22***
Conscientiousness	109.6	25.1	134.0	22.9	9.7	0.45***
SDE	77.1	12.4	88.4	12.8	8.7	0.41***
IM	73.7	19.6	87.8	17.9	7.2	0.34***
SDI	0.06	0.04	0.11	0.04	10.6	0.47***

Note: r_{PB} = point-biserial correlation between the condition indicator (honest and applicant conditions coded as 0 and 1, respectively) and personality and social desirability scales.
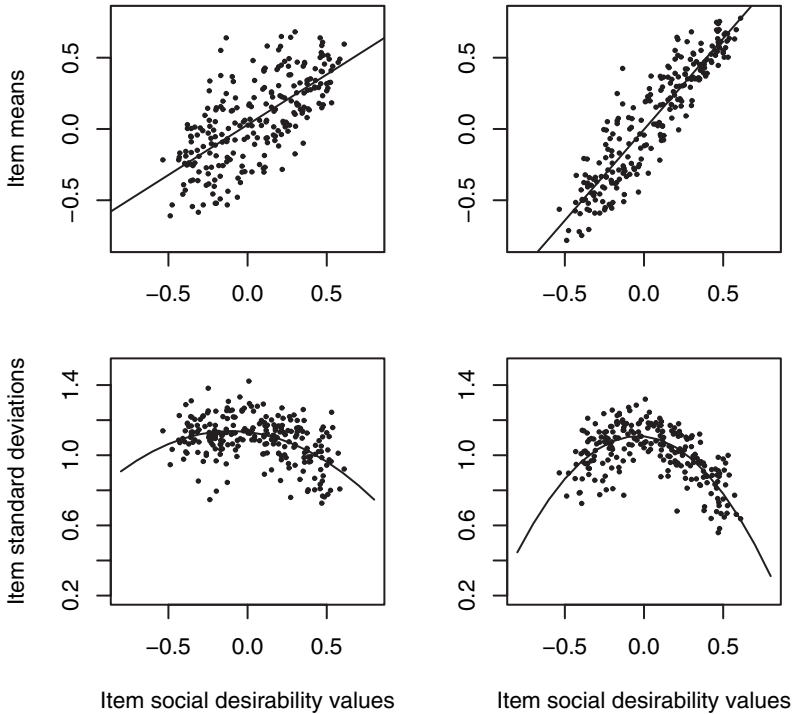***$p < 0.001$.

Figure 1. Regression of item means and standard deviation on their social desirability values. Left panel: 'honest' condition; right panel: 'applicant' condition.

honest and applicant conditions. However, as we expected, in a high-motivation condition, the item responses are more highly influenced by social desirability and a higher correlation between item means and social desirability values in the job applicant than in the honest condition was observed. The respective correlations were 0.91 (right panel) and
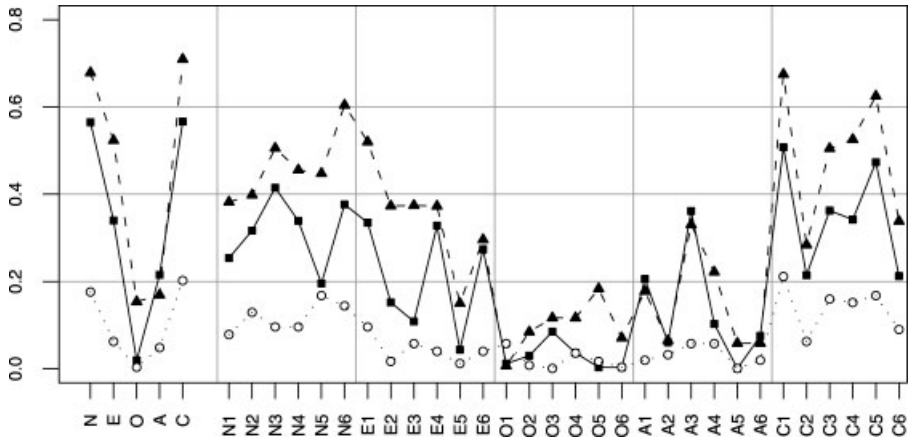
Figure 2.   Squared correlations of the NEO-PI-R domain and facet scale scores with SDI and condition. *Note:* Circles = difference between honest and job applicant conditions (squared point-biserial correlations); filled squares = squared correlations with social desirability index (SDI) in the honest condition; filled triangles = squared correlations with SDI in the job applicant condition.

0.68 (left panel). The correlation between item means in single peer-ratings and item social desirability values (not shown in the figure) was 0.86. The item standard deviations were significantly lower in the job applicant condition (mean = 1.01) than in the honest (mean = 1.09) condition: $t(239) = 12.7$; $p < 0.001$. In addition, the item standard deviations in the job applicant condition were better predicted from the respective social desirability values than in the honest condition; the R-squares from the quadratic model were 0.49 and 0.20 (see also Figure 2, lower panels). In peer-ratings, the average item standard deviation was 1.04 and the respective R-square was 0.20.

Social desirability component in the scale scores

The scale means in the honest and job applicant conditions, as well as the correlations of the scales with the social desirability index (SDI) are shown in Table 2. The mean differences expressed as point-biserial correlations (last column in Table 2) were in most cases smaller in absolute value than the respective correlations with SDI, but the effect sizes followed the same general pattern (see Figure 2). The correlation over the 30 facet scales of the NEO-PI-R between the squared mean-difference correlations and the squared correlations with SDI was 0.79 in the honest condition, and 0.86 in the job applicant condition. In addition, the correlation between the vectors of squared correlations with SDI in the two conditions was 0.91.

Detection of faking with social desirability indices

For investigating the detectability of faking, logistic regression analyses were performed predicting the condition indicator (the honest and applicant conditions were coded as 0 and 1, respectively) from the standardized social desirability indices. SDI alone predicted the condition indicator reasonably well (odds ratio [OR] = 3.54; $p < 0.001$); both of the BIDR scales also showed a significant relationship (ORs 2.34 for SDE and 1.70 for IM; both

ps < 0.001). In a model containing all three predictors, the odds ratios were 2.34 for SDI ($P < 0.001$), 1.57 for SDE ($P < 0.01$) and 1.31 for IM ($P = 0.07$).

Social desirability and consensual validity

In the honest condition, there was significant self-peer agreement on all the social desirability indices used in the present study. The self-other correlations were 0.44 for SDI, 0.21 for SDE and 0.40 for IM. In the job applicant condition, agreement was marginally significant for the SDI ($r = 0.16$, $p = 0.052$), and significant at the 0.05 level for the BIDR scales (rs $= 0.20$, and 0.24 for SDE and IM, respectively).

In all cases, self-other agreement was lower in the job applicant than in the honest condition (Table 3). The difference was significant ($p < 0.01$, one-sided) for Neuroticism and Conscientiousness, and marginally significant ($p < 0.1$) for Openness and Agreeableness. The bipartial correlations controlling for SDI in both self- and peer-ratings were, in most cases (except for Openness in the applicant condition self-ratings), higher than the corresponding zero-order correlations; this was true for both self-peer and peer-peer agreement. The significance of this general pattern of results can be tested by a binomial test. If SDR is unrelated to the consensual validity of personality scales, then controlling for it should lower the validity in at least a half of the cases, and probably more, considering the effect of partialling on the reliability of the scales. Therefore, a conservative null hypothesis in this case is that the probability of 'success' is 0.5; by binomial test, the probability of 14 'successes' in 15 trials is less than 0.001.

Table 3. Self-peer and peer-peer agreement: zero-order correlations and correlations controlling for social desirability indices.

Condition		r	r^a	r^b	r^c	r^d	Z^e	p	p^f
Honest	N	0.43	0.29	0.37	0.53	0.55	−1.8	0.072	0.042
	E	0.68	0.65	0.65	0.73	0.74	−1.6	0.108	0.044
	O	0.67	0.66	0.67	0.69	0.68	−1.8	0.074	0.024
	A	0.50	0.48	0.45	0.55	0.56	−1.4	0.166	0.096
	C	0.53	0.46	0.40	0.63	0.62	−1.8	0.072	0.032
Job applicants	N	0.11	0.17	0.05	0.25	0.20	−1.7	0.088	0.074
	E	0.62	0.60	0.65	0.68	0.71	−1.3	0.210	0.126
	O	0.56	0.55	0.56	0.56	0.54	0.2	0.810	1.000
	A	0.38	0.44	0.43	0.56	0.57	−3.6	0.000	0.000
	C	0.30	0.37	0.27	0.49	0.50	−2.3	0.022	0.012
Peer-peer agreement	N	0.33			0.45	0.45	−2.5	0.014	0.008
	E	0.51			0.61	0.60	−3.5	0.000	0.000
	O	0.36			0.40	0.40	−1.8	0.066	0.044
	A	0.43			0.47	0.48	−1.1	0.294	0.220
	C	0.40			0.43	0.44	−0.5	0.592	0.496

Note: The number of cases varied from 206 to 211 in the honest condition, from 148 to 150 in the job applicant condition, and from 289 to 292 for peer-ratings.
[a]Semipartial correlation controlling for the SDI in self-ratings only;
[b]Semipartial correlation controlling for the SDI in peer-ratings only;
[c]Bipartial correlation controlling for the SDIs in both self- and peer-ratings;
[d]Correlation between the scales computed after within-subject partialling of item desirabilities (Saucier, 1994);
[e]$\bar{Z}^*_{2-z}$ statistic for difference between two dependent correlations (Steiger, 1980; Meng, Rosenthal, & Rubin, 1992).
[f]p-value from permutation test (see text for further details).

To test the differences between the individual zero-order and bipartial correlations, the $\overline{Z}_2^*$ statistic proposed by Steiger (1980) was used; in addition to the two-sided p-value based on quantiles of the normal distribution (column labelled 'p' in Table 3), we also conducted a permutation test with 10 000 random permutations of the SDI. The last column in Table 3 reports the analog of a two-sided p-value from the permutation test (the probability that the $\overline{Z}_2^*$ statistic obtained from the permuted data was, in absolute value, smaller than the actual value). Both of these significance probabilities are about the equality of the zero-order and bipartial correlations, but in the permutation test, the effect of partialling on reliability is taken into account.[1] By the second criterion, more than a half of the differences between zero-order and bipartial correlations are significant, most notably, for self-peer agreement on Conscientiousness in both conditions, and Agreeableness in the applicant condition, and for the peer-peer agreement on Neuroticism and Extraversion. It can also be noted that in the honest condition, the effects of partialling (as measured by $\overline{Z}_2^*$) were rather homogenous across traits, whereas in the applicants condition and in peer ratings, there were larger effects for two traits (respectively Agreeableness and Conscientiousness, and Neuroticism and Extraversion).

Semipartial correlations (where the respective SDI was partialled out only from either self- or peer-reports) were always lower than the corresponding bipartial correlations and in a majority of cases also lower than the corresponding zero-order correlations. Controlling for the BIDR scales in self-reports generally either reduced or did not change the consensual validity. When controlling only for IM, there was an increase in Neuroticism (corrected $rs = 0.48$ in honest condition, 0.29 in applicants condition, and 0.38 in peer-ratings); the difference was significant (not controlling for multiple comparisons) only in the applicant condition. For the other scales, controlling for any of the BIR scales either had no noticeable effect on validity, or reduced it.

Moderator analyses

A series of moderator analyses was conducted predicting peer-rated traits from the self-ratings, social desirability indices, and their interactions. All variables were standardized before analyses. The SDI of self-ratings was a significant moderator in none of the analyses. In the honest condition, the SDI of peer-ratings moderated the self-peer agreement on Extraversion and Neuroticism; the results of these moderated regression analyses are shown in Table 4. In both cases, the change in R-square associated with the addition of the interaction component was between 1 and 2 per cent. The increase in SDI of peer-ratings by one standard deviation would produce a decrease of 0.12 in the standardized regression coefficient for Neuroticism, and a decrease of 0.11 in that of Extraversion, controlling in both cases for the SDI of self-ratings. The sign of the interaction effect was also negative but not significant for A and C; none of the interaction terms were significant in the applicant condition self-reports.

DISCUSSION

Operationalization of social desirability

One of the conclusions that can be drawn from the present results is that the measure of SDR proposed by Hofstee (2003), here called SDI, has several benefits over the traditional

---

[1] In other words, the second p-value is smaller because partialling out a random variable which is not a suppressor would, on the average, *lower* the self–other agreement rather than leave it intact.

Table 4. Moderated regression analyses: peer-rated neuroticism and extraversion regressed on self-ratings and the social desirability indices

	Neuroticism				Extraversion			
	b	SE	t	p	b	SE	t	p
Intercept	−0.01	0.04	−0.3	0.76	0.03	0.04	0.6	0.55
SDI_S	0.47	0.07	6.5	<0.0001	−0.37	0.06	−6.7	<0.0001
SDI_P	−0.71	0.05	−14.6	<0.0001	0.42	0.05	9.0	<0.0001
N	0.62	0.07	9.2	<0.0001	0.79	0.05	15.5	<0.0001
$SDI_P \times N$	−0.12	0.04	−2.9	0.004	−0.11	0.04	−2.7	0.008
R^2	0.61				0.64			
ΔR^2	0.016				0.012			

Note: The variables were standardized before analyses.
SDI_S = SDI in self-reports; SDI_P = SDI in peer-reports; $\triangle R^2$ = change in R^2 associated with the addition of interaction term.

measures; its effect on consensual validity is roughly equivalent to that obtained by Saucier's (1994) method of within-subject partialling of item social desirability values. Faking on personality scales was somewhat better detected by the SDI than by the two BIDR scales jointly.

An advantage of the SDI is its transparent logic: it measures directly the degree to which a respondent consistently agrees with socially desirable items and disagrees with undesirable items; it rests only on the assumptions of correct measurement of item desirabilities, and a sufficiently heterogeneous item pool. On the other side, when interpreting the scores on the traditional SD scales as social desirability rather than self-claimed honesty or friendliness, we are presuming that people actually cannot be as honest and friendly as these items state. The question is: how do we know? Moreover, we cannot be sure about how people understand these items. Take a typical SD item, 'I have never told a lie'. When a respondent agrees with that item, she most probably does not mean to agree with its Gricean 'timeless meaning' (Grice, 1957): all existent people have told at least a few small lies in their lives, and they know it. So in an ordinary conversation, the question 'have you never told a lie?' can hardly be seriously asked; within a supposedly serious test, its meaning may be implicitly transformed by some of the respondents into 'are you the kind of person who usually does not tell lies?' Another group of respondents may take the item literally but think that if they 'completely disagree' (which would be logically correct), they might be thought of as pathological liars. There are thus a variety of cognitive processes that can lead to any possible response option being selected; only some of these are concordant with the social desirability interpretation of the scores.

On the other hand, the large percentage of variance (more than 50%) that SDI shares with both Neuroticism and Conscientiousness may give rise to a concern about its interpretation as a measure of socially desirable responding. As the N and C items in the NEO PI-R have, on the average, more extreme social desirability values than items in the other scales (particularly Openness), SDI may be confounded with true Neuroticism and Conscientiousness. It should be noticed, however, that high correlations alone do not speak directly to this issue. Indirectly, the present results imply that the personality scores are more heavily confounded with SDR than our measure of SDR is confounded with substantive personality variance. First, self-other agreement on the personality scales was substantially higher than the self-other correlation on SDI (except in the case of

Neuroticism for which the uncorrected self-other correlation was approximately equal to that of the SDI). Second, removing the SDI-related variance improved self-other agreement on the personality scales; this result would be difficult to explain if SDI were considered just a blend of personality measures. The high correlations between SDI and the personality scale scores might look troubling because one is left with the impression that nothing remains of the true variance. This impression is misleading: even in the 'worst' case (Conscientiousness in the applicant condition), the correlation of the uncorrected scores with the corrected scores is higher than 0.50, as can be computed using the formula

$$r_{x,x \cdot z} = (1 - r_{x,z}{}^2)^{1/2}$$

The possible reasons for self-other correlation on the SDI are twofold: contamination resulting from the mean differences in the social desirability of different scales, and relationship effects (raters were selected by the participants, and those with higher SDI may have chosen the raters who liked them more). Conceptually, a method to eliminate confounding from the SDI would be to construct an inventory where the average social desirability values of each scale would be equal; in most inventories, this would entail multiple revisions of most of the items. This task is more complicated than it appears and may require changes in the content of the scales: some traits intuitively seem more important than others (for instance, 'ambitious' is judged to be more important than 'moderate' [Williams, Munick, Saiz, & FormyDuval, 1995]), and the judged importance and desirability are highly correlated (Williams, Satterwhite, & Saiz, 1998).

Evaluation in personality ratings: Beyond 'substance vs. style'

A second major conclusion from the present study is that socially desirable responding is, at least for some traits and in some situations, a suppressor of consensual validity—in other words, personality scale scores can be 'corrected' using SDI, and these corrected scores show higher consensual validity. As for moderator effects, the instructional manipulation (but not the SDI of self-ratings) moderated the consensual validity: self-other correlations were lower in the applicant condition, and in case of Neuroticism and Conscientiousness, these differences were not totally swept out by controlling for the SDI. Consequently, of the two perspectives on SDR described in the introduction, the bias or response shift model seems to hold for the inter-individual differences, whereas both bias and loss of information perspectives may be true for the intra-individual differences (e.g. motivation).

The present results do not show that SDI is a purely stylistic variable. If a response style is defined as 'a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content' (Paulhus, 1991), socially desirable responding is probably not such a style. On the contrary, social desirability is an important facet of item content even in the most trivial items. There may be real differences between persons who are very similar on all personality trait scores but have different values of SDI; such possibility is suggested, for example, by significant (but in the present study, moderate in size) self-other correlations on the social desirability measures (see also Hofstee, 2003; Hofstee & Hendriks, 1998). A sensible interpretation of the SDI is that it reflects the rater's general evaluation of the target (cf. Saucier, 1994). Evaluation can be highly consensual, but it is also likely to be more subjective, and less readily observable. Self-report SDI might be expected to be highly correlated with self-esteem; peer-report SDI would probably be related to a favorable attitude toward the target person.

Self- and other-enhancement

A further conclusion that the present results suggest is that socially desirable responding is not an exclusive property of self-reports. In fact, for an effect on consensual validity, the SDI had to be controlled in both self- and peer-reports; in addition, partialling out the SDI improved both self-peer and peer-peer agreement. An obvious technical explanation is that there was also interjudge agreement on SDI, which contributed to the interjudge agreement on personality scales; agreement on personality scales was thus both partly mediated *and* suppressed by social desirability.

In moderator analyses, a somewhat surprising result was that the self-report SDI was a significant moderator of agreement in none of the analyses, whereas the peer-report SDI was in two of them (for Extraversion and Neuroticism in the honest condition). Apparently, we are better judges of certain personality traits when we have a neutral rather than high or low opinion of the target person; this relationship does not hold, though, when we rate ourselves' personalities. This finding may be of some importance for research on social perception, but more research is required before any firm conclusions can be drawn from it.

Conclusions and Implications

In sum, it was shown in the present study that there are individual differences in the degree to which people's responses to personality items correspond to the social desirability values of these items. In a loose sense, these individual differences may be legitimately called a response style because removing this large portion of variance from the personality scales either improved or had no effect on validity. This result is in contrast with previous studies looking for a suppressor effect in self-other agreement, almost all of which have arrived at an opposite conclusion. Arguably, this divergence results from a more straightforward measure of SDR used in the present study, and from the recognition that peer-reports may also be biased by the SDR.

More broadly speaking, the present results suggest that when describing a target's personality characteristics, people tend to agree more on descriptive than evaluative aspects.[2] This is in line with some findings in the applied psychological literature where rater bias ('halo error') has been found to correlate negatively with accuracy (Fisicaro, 1988; Kiker & Motowidlo, 1998). Similarly, Viswesvaran, Schmidt, and Ones (2005, p. 111) have presented a statistical model of how idiosyncratic rater effects and rater by target interaction effects reduce agreement between raters.

While correcting for the SDI may remove some valid variance from the personality scales, in the present study it did more good than harm to the consensual validity. This does not automatically mean that SDI should always be partialled from the personality scales in applied contexts: before this could be recommended, additional research needs to be done with different samples of participants, and particularly with different criteria of validity. However, computing the SDI would be useful in most criterion validity studies using multidimensional personality questionnaires. Suppose, for instance, that we find a criterion to be related to all of the domains of the five-factor model, but to different degrees. Using the SDI, we might show that it is the common evaluative component (perhaps reflecting the true social desirability of a person's behaviours) of the personality scales that is related to the criterion; in that case, we might improve the prediction by using SDI instead of or in

---

[2]We thank an anonymous reviewer for suggesting this formulation.

addition to the personality scales. On the other hand, it may appear that these correlations are based on the descriptive component of the personality scales and are unchanged or heightened by controlling for the SDI. One might speculate that the latter case may be more typical when (a) the criterion involves human judgment and may be biased itself, and/or (b) when there is an obvious causal link between a particular trait and the criterion (e.g. sales performance of an insurance sales agent may be causally related to extraversion but not neuroticism; if this is true, one might expect that the criterion related validity of extraversion would be improved by controlling for social desirability, whereas that of neuroticism would be lowered). At this point, it is difficult to form any specific predictions for consensual validation studies, but it is likely that the effects of controlling for the SDI would depend on the relationship between target and rater, and on the importance of that relationship for the rater.

To end with, Jack Block (2003, p. 205) was probably too modest when he wrote concerning his 1965 book on social desirability, 'I had become a world expert on a topic in which no one subsequently was interested anymore'. Conversely, we believe to have shown that socially desirable responding (even if not a response set or style in the classical sense) is both of methodological importance and potentially of practical significance.

ACKNOWLEDGEMENTS

REFERENCES

Aavik, T., & Allik, J. (2002). The structure of Estonian personal values: A lexical approach. *European Journal of Personality, 16*, 221–235.

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI*. East Norwalk, CT: Appleton-Century-Crofts.

Block, J. (2003). My unexpected life. *Journal of Personality Assessment, 81*, 194–208.

Borkenau, P., & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality, 61*, 477–496.

Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6*, 199–214.

Caldwell-Andrews, A., Baer, R. A., & Berry, D. T. R. (2000). Effects of response sets on NEO PI-R scores and their relations to external criteria. *Journal of Personality Assessment, 74*, 472–488.

Cohen, J. (1982). Set correlation as a general multivariate data-analytic method. *Multivariate Behavioral Research, 17*, 301–341.

Comrey, A. L., & Backer, T. E. (1975). Detection of faking on the Comrey personality scales. *Multivariate Behavioral Research, 10*, 311–319.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI)*. Professional manual. Odessa, FL: Psychological Assessment Resources.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349–354.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational & Psychological Measurement, 23*, 699–720.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*, 90–93.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Holt, Rinehart & Winston.

Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*. New York: Holt, Rinehart & Winston.

Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology, 73*, 239–244.

Funder, D. C., & Colvin, C. R. (1997). Congruence of other's and self-judgments of personality. In R. R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego: Academic Press.

Grice, H. P. (1957). Meaning. *The Philosophical Review, 64*, 377–388.

Hofstee, W. K. B. (2003). Structures of personality traits. In I. B. Weiner, T. Millon, & M. Lerner (Eds.), *Handbook of Psychology: Personality and Social Psychology* (Vol. 5, pp. 231–254). Hoboken, NJ: Wiley.

Hofstee, W. K. B., & Hendriks, A. A. J. (1998). The use of scores anchored at the scale midpoint in reporting individuals' traits. *European Journal of Personality, 12*, 219–228.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521–551.

Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality, 14*, 265–278.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. C. (1992). Consensus at zero acquaintance: replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88–97.

Kiker, D. S., & Motowidlo, S. J. (1998). Effects of rating strategy on interdimensional variance, reliability, and validity of interview ratings. *Journal of Applied Psychology, 83*, 763–768.

McCrae, R. R., & Costa, P. T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882–888.

McCrae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood: A five-factor theory perspective*. (2nd ed.). New York: The Guilford Press.

Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175.

Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry, 10*, 23–34.

Nicholson, R. A., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist, 45*, 290–292.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.

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. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Mahwah NJ: Lawrence Erlbaum.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality & Social Psychology Bulletin, 21*, 100–108.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025–1060.

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

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality & Individual Differences, 37*, 1137–1152.

Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*, 263–275.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582–593.

Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35*, 463–482.

Saucier, G. (1994). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology, 68*, 141–154.

Saucier, G., Ostendorf, F., & Peabody, D. (2001). The non-evaluative circumplex of personality adjectives. *Journal of Personality, 69*, 537–582.

Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment, 68*, 127–138.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Smith, B. D., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*, 211–219.

Stöber, J., Dette, D. E., & Musch, J. (2002) Comparing continuous and dichotomous scoring of the balanced inventory of desirable responding. *Journal of Personality Assessment, 78*, 370–389.

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

Tesser, A., Pilkington, C. J., & McIntosh, W. D. (1989). Self-evaluation maintenance and the mediational role of emotion: The perception of friends and strangers. *Journal of Personality and Social Psychology, 57*, 442–456.

Topping, G. D., & O'Gorman, J. G. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences, 23*, 117–124.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance?: A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108–131.

Wiggins, J. S. (1959). Interrelationships among MMPI measures of dissimulation under standard and social desirability instruction. *Journal of Consulting Psychology, 23*, 419–427.

Wiggins, J. S. (1968). Personality structure. *Annual Review of Psychology, 19*, 293–350.

Williams, J. E., Munick, M. L., Saiz, J. L., & FormyDuval, D. L. (1995). Psychological importance of the 'Big Five': Impression formation and context effects. *Personality & Social Psychology Bulletin, 21*, 818–826.

Williams, J. E., Satterwhite, R. C., & Saiz, J. L. (1998). *The importance of psychological traits: A cross-cultural study*. New York: Plenum Press.