

Social Desirability in Spouse Ratings

**Vaka Vésteinsdóttir and
Eva D. Steingrimsdóttir**

Department of Psychology, University of Iceland,
Iceland

Adam Joinson

School of Management, University of Bath, UK

Ulf-Dietrich Reips

Department of Psychology, University of Konstanz, Germany

Fanney Thorsdóttir

Department of Psychology, University of Iceland, Iceland

Abstract

Whether or not socially desirable responding is a cause for concern in personality assessment has long been debated. For many researchers, McCrae and Costa laid the issue to rest when they showed that correcting for socially desirable responding in self-reports did not improve the agreement with spouse ratings on the Neuroticism, Extraversion, and Openness to Experience Personality Inventory. However, their findings rest on the assumption that observer ratings in general, and spouse ratings in particular, are an unbiased external criterion. If spouse ratings are also susceptible to socially desirable responding, correcting for the bias in self-rated measures cannot be assumed to increase agreement between self-reports and spouse ratings, and thus failure to do so should not be taken as evidence for the ineffectiveness of measuring and correcting for socially desirable responding. In the present study, McCrae and Costa's influential study was replicated with the exception of measuring socially desirable responding with the Marlowe–Crowne Social Desirability Scale, in both self-reports and spouse ratings. Analyses were based on responses from 70 couples who had lived together for at least one year. The results showed that both self-reports and spouse ratings are susceptible to socially desirable responding and thus McCrae and Costa's conclusion is drawn into question.

Corresponding Author:

Vaka Vésteinsdóttir, Haskoli Íslands, Aragata 14, Reykjavík 101, Iceland.
Email: vakav@hi.is

Keywords

Socially desirable responding, personality, response bias, spouse ratings, content overlap

Introduction

Socially desirable responding (SDR) generally refers to “the tendency to choose items that reflect socially approved behaviors” (Nunnally & Bernstein, 1994, p. 382). It is a socially shared belief of what will be approved of by others (e.g., Crowne & Marlowe, 1960) that is used to guide responding. SDR can therefore inflate scores on desirable items, and deflate scores on undesirable items. One approach to measuring SDR has been through the use of individually administered scales designed to capture this tendency (see Paulhus, 1991). When measured with such scales, SDR is seen as an individual difference variable—that is, respondents are presumed to have varying levels of the tendency to give socially desirable responses.

Researchers have raised concerns about SDR, in particular its nature and its effects, since the 1930s (e.g., Meehl & Hathaway, 1946; Rosenzweig, 1933; Vernon, 1934; Wiggins, 1968). For instance, there has been considerable debate on whether SDR is a cause for concern in personality assessment (e.g., Holden, 2007; Nevid, 1983; Rorer, 1965). SDR has traditionally been regarded as a response bias that poses a threat to the validity of personality scales (Cronbach, 1946; Edwards, 1953, 1957; Meehl & Hathaway, 1946). Under this assumption, an observed correlation between a personality scale and a scale designed to measure SDR indicates that the personality measures are confounded with error variance (Paulhus, 1991). Those who argue that SDR scales measure a substantive trait have challenged this interpretation and maintained that any observed correlation is the result of content overlap between the instruments. Taking this view, correlations between SDR scale scores and personality measures should thus not be regarded as indicators of response bias (Furnham, 1986; McCrae & Costa, 1983; Nicholson & Hogan, 1990; Smith & Ellingson, 2002).

More specifically, the debate centers on whether SDR scales measure a response style or a substantive personality trait, and whether SDR variance should be removed from personality scales. Because, in self-report studies, substance cannot be disentangled from response style, one strategy to resolve the debate is to compare self-reports with external criteria (Furnham, 1986), such as spouse ratings of a targets' personality. If SDR scales are indicators of a response style, then correcting for SDR should improve the agreement between self-reports and an external criterion, showing that SDR acted as a suppressor variable (e.g., Ganster, Hennessey, & Luthans, 1983).

In an influential study, McCrae and Costa (1983) used spouse ratings as an external criterion of personality traits in the domains of Neuroticism,

Extraversion, and Openness to Experience (NEO). Correcting for SDR in self-reports with the Marlowe–Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960) did not improve the agreement with spouse ratings, and for many traits lowered the agreement. Therefore, the authors concluded that SDR scales should be given substantive rather than artifactual interpretations.

Because correcting for SDR failed to improve agreement between self-reports and spouse ratings in the study of McCrae and Costa (1983), and numerous others who used observer ratings as external criteria (Borkenau & Ostendorf, 1992; Dicken, 1963; Kozma & Stones, 1987; Kurtz, Tarquini, & Iobst, 2008; McCrae et al., 1989; Pauls & Stemmler, 2003; Piedmont, McCrae, Riemann, & Angleitner, 2000), some researchers have concluded that SDR scales are not useful for enhancing the validity of personality measurements (e.g., Borkenau & Ostendorf, 1992) or even that researchers should not make an effort to detect SDR in personality measurements at all (e.g., Piedmont et al., 2000). Costa and McCrae (1997) intentionally exclude validity scales to identify SDR to the Revised NEO Personality Inventory (NEO-PI-R), claiming that they remain unconvinced of the utility of such scales, and, in support of that opinion, cite their own 1983 study among other studies using observer ratings. As Ones, Viswesvaran, and Reiss (1996) note, for many researchers the SDR issue is a “methodological dead horse” (Nevid, 1983, p. 139), laid to rest by McCrae and Costa (1983).

These conclusions are an overgeneralization from McCrae and Costa’s (1983) results and it should be noted that McCrae and Costa did not rule out the possibility of SDR. They acknowledged that their respondents were disinterested volunteers who answered anonymously and were not motivated to distort their answers. Therefore, SDR might still be a problem in situations where motivation to distort is high, for example, in personnel assessments (as has been demonstrated; e.g., Rosse, Stecher, Miller, & Levin, 1998). In addition, a number of other studies have shown that SDR degrades the validity of personality measures (e.g., Ellingson, Sackett, & Hough, 1999; Holden, 2007; Viswesvaran & Ones, 1999). Topping and O’Gorman (1997), for example, reported lowered self-other agreement on the NEO Five Factor Inventory scales under instructions to “fake good” in the self-reports. Furthermore, SDR scales can detect faking (Holden, 2007; Lambert, Arbuckle, & Holden, 2016).

Previous objections notwithstanding, McCrae and Costa’s (1983) findings rest on the assumption that observer ratings in general, and spouse ratings in particular, are an unbiased external criterion. More specifically, McCrae and Costa note that although both self-reports and spouse ratings of personality may contain social desirability (abbreviated SD in their discussion) . . .

There is no reason to suspect that the SD of the subject would influence his or her spouse’s ratings. The only variance common to the two sources would be real trait variance, and this alone could account for the correlation between them. (p. 884)

This assumption is questionable because if both partners are responding in a socially desirable manner, their responses are bound to have more in common than if one of them did and not the other, and therefore the common variance of the two sources may not just be real trait variance. It seems intuitive that if spouses have a tendency to respond in a socially desirable manner, they will see it desirable to have a partner with socially desirable qualities and rate them accordingly.

For instance, Funder and Colvin (1997, p. 625) suggest that the “self”-enhancement bias may be poorly named because enhancement effects have also been found when comparing acquaintance ratings to those made by strangers. These findings may be explained by the Self-Evaluation Maintenance Model, which assumes that peoples’ self-evaluations are partly determined by a process whereby the good qualities of close others are perceived as reflecting something about oneself (Tesser, Pilkington, & McIntosh, 1989).

There are further reasons to believe that spouse ratings may be influenced by SDR. SDR is based on shared beliefs about what will be approved of by others (e.g., Crowne & Marlowe, 1960). Responding in a socially desirable manner is thus responding in accordance with social norms, or more specifically injunctive norms, which specify what people ought to do, that is, what will be approved or disapproved of by others, and the anticipated social sanctions for not acting in accordance with the norm (Cialdini, Reno, & Kallgren, 1990; Reno, Cialdini, & Kallgren, 1993). Hence, social desirability is very different from personal preferences because the direction of the resulting bias (the bias produced by presenting an image that will be viewed favorable by others) is predictable through the knowledge of injunctive norms. In fact, peoples’ judgments of how desirable or undesirable a scale item is are highly consistent, even when obtained from different groups of people (Edwards, 1957).

Because of these shared beliefs, SDR produces similarity between measurements contaminated with SDR, not only when produced by the same individual but also between individuals. McCrae and Costa (1983) express the same logic in their discussion of the expected consequences of SDR, but for some reason they implicitly take it to apply only to measurements obtained from the same person: “. . . such systematic distortion would be seen in the correlations between scales: Measures of traits that are socially desirable may covary, not because they are substantively related but because they are both susceptible to the operation of SD biases” (p. 883). The social aspect of SDR is, however, what makes SDR function in the same way for people in general. It follows that, if both self-reports and spouse ratings are contaminated with SDR, the correlation between the two would be strengthened, and thus “partialing out the variance in self-reports that is due to SD” (p. 884), as was done by McCrae and Costa, could not be expected to increase the correlation. Therefore, the failure of this method to increase the correlation between self-reports and spouse ratings cannot be taken as evidence for the ineffectiveness of measuring and correcting for SDR.

Because of the great influence McCrae and Costa's (1983) study had, and still has, on the interpretation of SDR, the purpose of the present study is to replicate it using the NEO-PI-R, with the exceptions of obtaining measures of spouses' SDR, to test McCrae and Costa's assumption that spouse ratings of personality are an unbiased external criterion of their partners' personality. A measure of spouses' SDR cannot be said to share content with a rating of their partner's personality as these are not measures of the same individual. Thus, if spouses' SDR correlates with spouse ratings of personality this can be taken as an indication of response distortion.

Furthermore, because SDR is a social construct, based on shared beliefs of what is desirable, spouses' SDR is hypothesized to correlate with spouse ratings of personality in a similar way as subjects' SDR correlates with self-reports of personality. McCrae and Costa (1983) used the Neuroticism–Extraversion–Openness (NEO) Inventory as a measure of personality. In their study, the correlation of these three traits with the measure of SDR (MCSDS) were: $r = -.49$ ($p < .001$) with Neuroticism (N), $r = .15$ ($p < .05$) with Extraversion (E) and $r = -.13$ (insignificant) with Openness (O). A similar pattern as observed by McCrae and Costa is expected in the present study for both partners.

N is an undesirable trait and thus a significant negative correlation is expected. E is a somewhat desirable trait, though introversion (the other end of the continuum) is not necessarily undesirable. This might explain why some studies have not found an association between SDR and E (see, e.g., Ones et al., 1996 for a meta-analysis of the association between personality and SDR), whereas McCrae and Costa (1983) and other more recent studies have found a mild positive correlation between SDR (MCSDS) and E (e.g., Holden & Passey, 2010; Kurtz et al., 2008). It is therefore not entirely clear whether to expect SDR to be uncorrelated, or have a mild positive correlation, with E, although the latter is more likely since we intend to use the same measures of personality and SDR as McCrae and Costa. O, however, is neither desirable nor undesirable as reversed items have a reference to rational thinking, which cannot be conceived of as undesirable, and thus there is no reason to expect a correlation between O and the MCSDS. In addition, because McCrae and Costa (1983) suggest that “individuals high in SD will appear to score higher on measures of adjustment, conscientiousness, agreeableness and other socially desirable traits than they actually are” (p. 883), the traits Conscientiousness (C) and Agreeableness (A) were added in the present study, with the expectation of a positive correlation to the MCSDS for both partners.

Method

Participants

A convenience sample of 70 heterosexual couples that had lived together for at least one year participated in the study. Couples were recruited by asking

people (most of whom worked at the same firm) to participate by giving a self-evaluation of their personality and asking their partner to rate them as well. An approximately equal number of men and women were recruited for participation. The recruitment process was non-random and focused on obtaining a relatively equal number of men and women for completion of self-reports, and thus observer reports. Those who agreed to participate were given verbal instructions to give a self-report of their own personality and ask their partner to fill out a similar questionnaire containing an observer report of personality. Two of the initially recruited 72 couples were excluded after initial screening of participants, as they did not meet the requirements of the study (returning a blank spouse-rating questionnaire and neither partner giving a valid response to the length of cohabitation question). Self-reports were obtained from 34 men (48.6%) and 36 women (51.4%), with corresponding spouse ratings from 34 women and 36 men. Participants' age ranged from 21 to 59 ($M = 33$, $SD = 9.7$). All participants were volunteers and did not receive compensation for study participation.

The average length of cohabitation was nine years and nine months (ranging from 1 year and two months to 39 and a half years). There was, however, some inconsistency in the reported length of cohabitation between some of the participants and their partners. In three cases, only one of the partners gave a response to the question and in 14 cases, the responses from the two partners did not match (and were thus averaged for the above calculation of average length of cohabitation). In 12 of those 14 cases, the discrepancy can be considered relatively minor (four months on average). For the remaining two couples, the difference in reporting was three and five years (both couples were however married, and both reported the exact same number of years married). A little under half of the participating couples in the study were married (41.1%) and over two-thirds had one or more children in their care (70%).

Measures

Personality measure. The NEO-PI-R inventory is a 240-item measure of the Big Five personality traits: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Each dimension has six facets consisting of eight items rated on a five-point scale ranging from "strongly disagree" to "strongly agree" (Costa and McCrae, 1992). The Icelandic version of the self-report form (Form S) and the observer rating form (Form R) (Jónsson & Bergórsson, 2004) were used.

Neuroticism (N) contains 48 items that measure the predisposition to experience psychological distress. For the N domain, reported internal consistency of the Icelandic self-report form is 0.91 (Jónsson & Bergórsson, 2004), which is almost the same as for the original version (0.92; Costa & McCrae, 1992).

Internal consistency of the Icelandic observer rating form (0.90) is also almost the same as for the original version (0.91; McCrae & Terracciano, 2005). In the present study, Cronbach's alpha values for N were 0.93 (Form S) and 0.95 (Form R).

Extraversion (E) contains 48 items used to measure, for example, friendliness and preference for the company of others. For the E domain, reported internal consistency of the Icelandic self-report form is 0.88 (Jónsson & Bergórsson, 2004), almost the same as the original version (0.89; Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form is equal to the original version (0.91; McCrae & Terracciano, 2005). In the present study, Cronbach's alpha values for E were 0.92 (Form S) and 0.93 (Form R).

Openness to Experience (O) contains 48 items used to measure, for example, aesthetic sensitivity and intellectual curiosity. For the O domain, reported internal consistency of the Icelandic self-report form is 0.87 (Jónsson & Bergórsson, 2004), which is equal to the original version (Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form (0.88) is also equal to the original version (McCrae & Terracciano, 2005). In the present study, Cronbach's alpha values for O were 0.81 (Form S) and 0.86 (Form R).

Agreeableness (A) contains 48 items used to measure, for example, helpfulness and cooperation with others. For the A domain, reported internal consistency of the Icelandic self-report form is 0.82 (Jónsson & Bergórsson, 2004), which is similar to the original version (0.86; Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form (0.92) is almost equal to the original version (0.93; McCrae & Terracciano, 2005). In the present study, Cronbach's alpha values for A were 0.85 (Form S) and 0.90 (Form R).

Conscientiousness (C) contains 48 items used to measure, for example, organizing and carrying out tasks. For the C domain, reported internal consistency of the Icelandic self-report form is 0.88 (Jónsson & Bergórsson, 2004), which is similar to the original version (0.90; Costa & McCrae, 1992). Internal consistency of the Icelandic observer rating form (0.95) is almost equal to the original version (0.94; McCrae & Terracciano, 2005). In the present study, Cronbach's alpha values for C were 0.91 (Form S) and 0.94 (Form R).

Social desirability measure. An Icelandic translation of the MCSDS was used as a measure of SDR (Vésteinsdóttir, Reips, Joinson, & Thorsdóttir, 2012). The MCSDS is a true–false questionnaire consisting of 33 statements of either uncommon but socially desirable behavior or common but undesirable behavior (Crowne & Marlowe, 1960). All participants (both subjects and spouses) were instructed to rate themselves on the MCSDS. In the present study, Cronbach's alpha values were 0.77 and 0.75 for the subjects' self-reports and for the spouses' self-reports, respectively.

Procedure

Each couple received an envelope containing the self-report (Form S) and observer rating form (Form R) of the NEO-PI-R inventory. Attached to both Form S and Form R was the MCSDS, which both partners were specifically asked to give self-reports on. Participants filled out the questionnaires in their own homes under the instructions to answer one of the two forms independently of their spouses. Upon completion, participants returned the questionnaires in an envelope labeled only with identification numbers to ensure complete anonymity.

Results and Discussion

Descriptive statistics for scales used in this study are shown in Table 1.

The mean scores and standard deviations (SD) of the MCSDS were in line with those reported in previous studies (for an overview of descriptive statistics from previous studies, see Vésteinsdóttir, Reips, Joinson, & Thorsdottir, 2015). No difference was observed between mean MCSDS scores for subjects and spouses ($t(130) = .056, p = .95$) and no gender differences were observed. For the male subjects ($n = 31$), the mean score ($M = 15.7, SD = 6.0$) on the MCSDS did not differ from female subjects' ($n = 34$) mean score ($M = 16.5, SD = 4.1$), $t(52.52) = -.608, p = .55$. The male spouses' ($n = 34$) mean score ($M = 16.5, SD = 5.7$) also did not differ from the female spouses' ($n = 33$) mean score ($M = 15.7, SD = 4.2$), $t(60.52) = .613, p = .54$.

Table 1. Descriptive statistics for the five dimensions of the NEO-PI-R and the MCSDS.

		N	Mean	SD	Minimum	Maximum
MCSDS	Subject	65	16.2	5.1	2	30
	Spouse	67	16.1	5.0	7	30
Neuroticism	Subject	66	90.9	25.2	52	155
	Spouse	66	83.9	29.7	24	154
Extraversion	Subject	64	118.0	22.0	60	161
	Spouse	62	118.5	25.1	57	168
Openness	Subject	63	109.6	14.9	76	150
	Spouse	67	105.0	17.8	68	161
Agreeableness	Subject	64	122.6	16.0	76	152
	Spouse	62	126.7	20.8	89	172
Conscientiousness	Subject	66	116.0	21.4	60	165
	Spouse	64	123.4	26.9	55	170

MCSDS: Marlowe–Crowne Social Desirability Scale; NEO-PI-R: Revised NEO Personality Inventory.

Table 2 shows the correlation between subjects' measured SDR (MCSDS_{subject}) and self-ratings of the five dimensions of the NEO-PI-R: Neuroticism, Extraversion, Openness to Experience, Conscientiousness, and Agreeableness, together with spouse's SDR (MCSDS_{spouse}) and spouse ratings of their partners' personality traits.

The first line in Table 2 shows the correlations between subjects' self-ratings of personality and SDR. The results are mostly in line with what previous studies on the relationship between self-rated personality and SDR have found (Holden & Passey, 2010; Kurtz et al., 2008; McCrae & Costa, 1983). However, contrary to our predictions, the correlation between E and the SDR did not reach significance in the current study. This may be attributed to a smaller sample than used in previous studies as the correlation coefficient for E and MCSDS is in line with that reported in previous studies ($r = .22, p = .05$; $r = .10, p < .01$; $r = .15, p < .05$, respectively). Overall the pattern of results is very similar between the studies. Kurtz et al. (2008) and McCrae and Costa (1983) interpreted this same pattern of correlations between the SDR and personality traits as content overlap, and Holden and Passey (2010) as method variance attributable to self-report.

However, when looking at the observed correlations between SDR and spouse ratings (shown in the second line of Table 2), it can be seen that spouses' SDR correlates with their ratings of partners' personality in the expected direction and in a similar pattern as their partners' SDR and self-reported personality. Although the correlation between subjects' SDR and A is somewhat stronger than between spouses' SDR and A, the difference between the two correlation coefficients for SDR and A is insignificant, $p = .14$, (however, due to the small sample size of the current study and the resulting lack of power in significance testing, this finding should be interpreted with caution). On the whole, similar results were obtained for both partners. As observed for subjects', a negative correlation was found for spouses' between SDR and the undesirable trait N, and positive correlations between SDR and the desirable traits A and C (but not for the somewhat desirable trait E).

Table 2. Correlation between self-reports of personality and subjects' SDR, together with the correlation between spouse ratings of personality and spouses' SDR.

	Neuroticism		Extraversion		Openness		Agreeableness		Conscientiousness	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
MCSDS _{subject}	-.38**	<.01	.19	.07	.02	.43	.48**	<.01	.29*	.01
MCSDS _{spouse}	-.27*	.02	.12	.19	.16	.10	.24*	.04	.33**	.01

Note: Due to the small sample size, a pairwise deletion was used in all calculations to maintain power. *Correlation significant at the .05 level (one-tailed). **Significant at the .01 level (one-tailed).

This finding can neither be interpreted as content overlap nor method variance attributed to self-report. Spouses' SDR scores cannot be taken to indicate content of their partners' personality. Nor can method variance, attributed to self-report, account for the correlation between self-reports (the MCSDS) and rater-reports (spouse ratings of personality). Both subject and spouse ratings of personality are associated with SDR in a similar way, and because the correlation between spouses' SDR and spouse ratings cannot be written off as content overlap, a more likely interpretation is that both measures are contaminated by SDR.

These findings mean that spouse ratings cannot be taken to be an unbiased external criterion. Partialing out the variance due to subjects' SDR cannot be expected to increase the correlation between self-reports and spouse ratings of personality, and thus the failure of this method to do so does not indicate that the MCSDS is an invalid measure of SDR.

However, it does not follow that partialing out the subjects' SDR should decrease the overall correlation between self-reports and spouse ratings, because SDR is seen here as an individual difference variable, that is, a tendency of the respondent, not necessarily exhibited by both partners. Thus, only in the situation where both respond in a socially desirable manner would partialing out the subjects' SDR decrease the correlation between self-reports and spouse ratings. If, however, one partner responds in a socially desirable manner and the other does not, partialing out the subjects' SDR could either have no effect or increase the correlation, depending on which partner exhibits the tendency. The third possibility is that neither partner displays this tendency, in which case partial correlation would have no effect.

As obvious as the result of the third possibility is, McCrae and Costa (1983) make no mention of this in relation to the insignificant correlation between O and the MCSDS in their study. Because the correlation was insignificant to begin with, correcting for MCSDS is a pointless procedure that adds nothing to the argument that MCSDS is an invalid measure of SDR, even if their assumption of spouse ratings being an unbiased criterion was correct. Much the same holds for E, as the correlation between E and the MCSDS was low ($r = .15$) and therefore correcting for the MCSDS cannot be assumed to have much of an effect. When correcting the correlation between self-reported and spouse rated N for the MCSDS, the correlation dropped from .58 to .49, which McCrae and Costa took to indicate content overlap. As explained above, this interpretation is unwarranted because spouses' SDR also correlates with N and thus correcting for MCSDS lowers the correlation for couples where both respond in a socially desirable manner. It is only when the subject responds in a socially desirable manner and the spouse does not that an increase in correlation can be expected when correcting for the subject's SDR.

A limitation to the present study is the small sample size and consequently the above-explained correction patterns cannot be demonstrated with the current data. However, with a sufficient sample size, future research could demonstrate

this by splitting the sample into three groups, one with couples where both have high scores on the MCSDS (HH group), a second group where one partner would have a high score on the MCSDS and the other would have a low score (HL group), and a third group where both would score low on the MCSDS (LL group). Correcting for the MCSDS in the HH group should reduce the correlation between self-reported and spouse-rated personality, whether the correction is made only for the subjects' SDR or both the subjects' and the spouses' SDR. In the second group correcting for the subjects' SDR could bring the correlation up or down depending on the composition of the sample, but correcting for both partners' SDR should increase the correlation. Corrections for the MCSDS should have no effect on correlation between self-reports and spouse ratings in the LL group; however, the correlation observed for this group could be taken to be unbiased by SDR and therefore the corrected correlation obtained for the other two groups should approximate the correlation in the LL group, given that the correlation is corrected for both partners' SDR.

Conclusion

The debate over the validity of measures of SDR is essentially a debate over the interpretation of correlational relationships between measures of SDR and personality, which centers on the content overlap argument proposed by McCrae and Costa (1983), among others. The argument follows that SDR scales are not a measure of SDR but a measure of substantive traits, and thence that any observed correlation between SDR scales and personality measures is therefore the result of content overlap between measures. McCrae and Costa used spouse ratings as an external criterion, claiming it to be unbiased. As the current study shows, spouse ratings are affected by SDR and are thus an invalid criterion. Furthermore, the change in correlation due to correction for SDR is dependent on the composition of the sample, that is, to what extent neither, just one, or both partners display SDR and to what degree, and cannot therefore be predicted with accuracy. In any case, spouse ratings of their partners' personality cannot be claimed to have content overlap with spouses' SDR, that is, spouses' SDR cannot be taken to be a measure of their partners' N, A, and C. With content overlap ruled out, the correlation between spouses' SDR and spouse ratings of personality can hardly be interpreted as displays of a substantive trait, unless giving favorable descriptions can be called a trait, which would just be a relabeling of the tendency to give desirable answers—otherwise known as socially desirable response style.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was part funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1) and The Eimskip Fund of The University of Iceland (Háskólasjóður Eimskipafélags Íslands). The funding sources had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

References

- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality*, *6*(3), 199–214. doi: 10.1002/per.2410060303
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. doi: 10.1037/0022-3514.58.6.1015
- 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.
- Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: The revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment*, *68*(1), 86–94. doi: 10.1207/s15327752jpa6801_7
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*(4), 475–494. doi: 10.1177/001316444600600405
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. doi: 10.1037/h0047358
- Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement*, *23*, 699–720. doi: 10.1177/001316446302300406
- 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*(2), 90–93. doi: 10.1037/h0058073
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155–166. doi: 10.1037/0021-9010.84.2.155
- 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, CA: Academic Press.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*, 385–400. doi: 10.1016/0191-8869(86)90014-0
- Ganster, D. C., Hennessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal*, *26*(2), 321–331. doi: 10.2307/255979

- Holden, R. R. (2007). Socially desirable responding does moderate personality scale validity both in experimental and in nonexperimental contexts. *Canadian Journal of Behavioural Science, 39*(3), 184–201. doi: 10.1037/cjbs2007015
- Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49*(5), 446–450. doi: 10.1016/j.paid.2010.04.015
- Jónsson, F. H., & Bergórsson, A. (2004). Fyrstu niðurstöður úr stöðlun NEO-PI-R á Íslandi [First results of NEO-PI-R standardization in Iceland]. *Sálfræðiritið, 9*, 9–16.
- Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: A systematic evaluation. *Journal of Gerontology, 42*, 56–59. doi: 10.1093/geronj/42.1.56
- Kurtz, J. E., Tarquini, S. J., & Iobst, E. A. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences, 45*(1), 22–27. doi: 10.1016/j.paid.2008.02.012
- Lambert, C. E., Arbuckle, S. A., & Holden, R. R. (2016). The Marlowe–Crowne social desirability scale outperforms the BIDR impression management scale for identifying fakers. *Journal of Research in Personality, 61*, 80–86. doi: 10.1016/j.jrp.2016.02.004
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*(6), 882–888. doi: 10.1037/0022-006X.51.6.882
- McCrae, R. R., Costa, P. T., Dahlstrom, W. G., Barefoot, J. C., Siegler, I. C., & Williams, R. B. (1989). A caution on the use of the MMPI K-correction in research on psychosomatic medicine. *Psychosomatic Medicine, 51*(1), 58–65. doi: 10.1097/00006842-198901000-00006
- McCrae, R. R., & Terracciano, A. (2005). Universal features of personality traits from the observer’s perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*(3), 547–561. doi: 10.1037/0022-3514.88.3.547
- 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*(5), 525–563. doi: 10.1037/h0053634
- Nevid, J. S. (1983). Hopelessness, social desirability, and construct validity. *Journal of Consulting and Clinical Psychology, 51*, 139–140. doi: 10.1037/0022-006X.51.1.139
- Nicholson, R. A., & Hogan, R. T. (1990). The construct validity of social desirability. *American Psychologist, 45*, 290–292. doi: 10.1037/0003-066X.45.2.290
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychological theory* (3rd ed.). New York, NY: McGraw-Hill.
- 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*(6), 660–679. doi: 10.1037/0021-9010.81.6.660
- 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* (pp. 17–59). San Diego, CA: Academic Press.
- Pauls, C. A., & Stemmler, G. (2003). Substance and bias in social desirability responding. *Personality and Individual Differences, 35*(2), 263–275. doi: 10.1016/S0191-8869(02)00187-3

- 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*(3), 582–593. doi: 10.1037/0022-3514.78.3.582
- Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology*, *64*(1), 104–112. doi: 10.1037/0022-3514.64.1.104
- Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin*, *63*(3), 129–156. doi: 10.1037/h0021888
- Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, *40*(4), 337–354. doi: 10.1037/h0074916
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634–644. doi: 10.1037/0021-9010.83.4.634
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology*, *87*(2), 211–219. doi: 10.1037/0021-9010.87.2.211
- Tesser, A., Pilkington, C. J., & McIntosh, W. D. (1989). Self-evaluation maintenance and the meditational role of emotion: The perception of friends and strangers. *Journal of Personality and Social Psychology*, *57*, 442–456. doi: 10.1037/0022-3514.57.3.442
- 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. doi: 10.1016/S0191-8869(97)00006-8
- Vésteinsdóttir, V., Reips, U.-D., Joinson, J., & Thorsdottir, F. (2012, November). *Psychometric properties of an Internet administered version of the Marlowe-Crowne Social Desirability Scale*. Paper presented at the Sixth ISM Workshop on Internet Survey Methodology, Ljubljana, Slovenia. Supported by COST Action IS1004. Retrieved from www.webdatanet.eu
- Vésteinsdóttir, V., Reips, U. D., Joinson, A., & Thorsdottir, F. (2015). Psychometric properties of measurements obtained with the Marlowe–Crowne Social Desirability Scale in an Icelandic probability based Internet sample. *Computers in Human Behavior*, *49*, 608–614. doi: 10.1016/j.chb.2015.03.044
- Vernon, P. E. (1934). The attitude of the subject in personality testing. *Journal of Applied Psychology*, *18*(2), 165–177. doi: 10.1037/h0074033
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality assessment. *Educational and Psychological Measurement*, *59*, 197–210. doi: 10.1177/00131649921969802
- Wiggins, J. S. (1968). Personality structure. *Annual Review of Psychology*, *19*(1), 293–350. doi: 10.1146/annurev.ps.19.020168.001453.

Author Biographies

Vaka Vésteinsdóttir is a PhD student in Methodology at the University of Iceland, Department of Psychology, but will be joining the Psychological Methods, Assessment and iScience team at the University of Konstanz as a postdoc. She obtained her BA and MA degrees at the University of Iceland

on the topics of telephone interviewers' voice and acceptance rates, and honesty testing. Her research interests are in survey methodology, particularly in questionnaire construction and measurement bias, with a focus socially desirable responding.

Eva D. Steingrimsdottir has a BS degree in Psychology from the University of Iceland. She also has a BS degree in Computer Science from the University of Iceland. She currently works as a software developer.

Adam Joinson holds the post of Professor of Information Systems at the University of Bath, School of Management. His research focuses on the interaction between psychology and technology, with a particular focus on how technology can be to shape behaviour, social relations and attitudes. Recently this work has taken in privacy attitudes and behaviours, the social impact of monitoring technology, computer-mediated communication and the human aspects of cyber security. The EPSRC, ESRC, EU, British Academy and UK Government have funded this work, and he has published over 80 articles in the field, as well as editing the *Oxford Handbook of Internet Psychology* (OUP, 2007) and authoring two books on psychology and technology. He is principal investigator for the Cyber-Security Across the LifeSpan (www.csalsa.uk) project and co-investigator of the Centre for Research and Evidence on Security Threats (www.crestresearch.ac.uk). His website is www.joinson.com.

Ulf-Dietrich Reips is a full professor in the Faculty of Sciences at the University of Konstanz, where he holds the Chair for Psychological Methods, Assessment, and iScience. For more than two decades he has been working on Internet-based research methodologies (or Internet science), the psychology of the Internet, measurement, development, the cognition of causality, personality, privacy, Social Media, crowdsourcing, and Big Data. In 1994, he founded the Web Experimental Psychology Lab, the first laboratory for conducting real experiments on the World Wide Web. Ulf was elected the first non-North American president of the Society for Computers in Psychology (SCiP) and he is the founding editor of the *International Journal of Internet Science*. Many of his over 140 publications are among the most highly cited in their journals, see <http://www.uni-konstanz.de/iscience/reips/pubs/publications.html>. Ulf has worked, lived, and studied in California, Colorado, Israel, Germany, Spain, Switzerland, and the UK. In 2014, he was ranked 7th of "Top Scientists working at Spanish Private Universities" by the Consejo Superior de Investigaciones Cientificas, Spain. Recently, he has been asked to direct the Leibniz institute for Psychology information in Trier, Germany. Ulf and his team develop and provide free Web tools for researchers, teachers, students, and the public. They received numerous awards for their Web applications (available from the

iScience Server at <http://iscience.eu/>) and methodological work serving the research community.

Fanney Thorsdottir is an associate professor of Psychology at the University of Iceland, School of Health and a director of the Center of Methodological Research. She has a BA and a Master (Psychology) from the University of Iceland, a Diploma in Social Science Data Analysis from the University of Essex and a PhD (psychology) from the London School of Economics. Her research interests are in the areas of psychometrics and survey methodology with a particular focus on measurement bias in self-report measurements.