Socially Desirable Responding Suppresses the Association Between Self-Assessed Intelligence and Task-Based Intelligence

Gilles E. Gignac

University of Western Australia

A version to be published in Intelligence (Accepted May 9th, 2018)

Author Note

Correspondence should be addressed to Gilles E. Gignac, School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia, 6009, Australia. E-mail: gilles.gignac@uwa.edu.au; the author(s) received no financial support for the research, authorship, and/or publication of this article; thanks to Courtney Weber, Jordan Lee, Sonia Sappl, and Ryan Ho Junxian for data collection.
Abstract

The response bias hypothesis specifies that the predictive capacity of a predictor should be enhanced by controlling for the effects of bias on the predictor variable, in particular, socially desirable responding (SDR) bias. To-date, the vast majority of the SDR research in the area, which is principally personality related, has failed to support the response bias hypothesis, as SDR suppressor effects have not been observed. Consequently, it has been contended that SDR is not a problem for self-report measurement, that SDR measures may themselves be indicators of trait variance, and that it was likely impossible to determine whether an elevated SDR score reflected a trait or response bias. However, in contrast to personality, intelligence is an area within which comparisons between subjective scores (self-reported) and objective scores (task-based) can be made. Consequently, the purpose of this investigation was to test the response bias hypothesis (N = 253) with self-report measures of intellectual and emotional intelligence (SRIQ and SREI) and task-based measures of intellectual and emotional intelligence (TBIQ and TBEI), in conjunction with a multi-dimensional measure of SDR (Balanced Inventory of Desirable Responding; BIDR). The percentage of variance accounted for in TBIQ by SRIQ, and in TBEI by SREI, increased by 1% and 2.1%, respectively, when SDR was included in the model. The 1% to 2.1% increases in criterion (concurrent) validity were interpreted as practically significant, based on previously published simulation work. Finally, it was concluded that self-report measures may be non-negligibly influenced by individual differences in SDR, and that the BIDR may possess some validity as an indicator of individual differences in socially desirable responding.

Keywords: self-assessed intelligence, emotional intelligence, response bias hypothesis, socially desirable responding
Socially Desirable Responding Suppresses the Effect Between Self-Assessed Intelligence and Task-Based Intelligence

Researchers and practitioners alike continue to express concerns about the possibility that self-report measures may be affected adversely by socially desirable responding (SDR; Tracey, 2016). However, the empirical literature in the area is decidedly mixed: Some have argued that SDR is a problem to be considered seriously (e.g., Holden, 2007), while others have contended that it is probably not (e.g., Spector, 2006). Even more fundamentally, it has been contended that typical measures of SDR are likely invalid indicators of socially desirable responding, as responders may actually possess the socially attractive qualities implied by their high SDR scores (McCrae & Costa, 1983). More recently, Paunonen and Lebel (2012) suggested that it was likely impossible to determine whether an elevated SDR score reflected a genuinely high level of the trait or an inflated score due to response bias.

Although Paunonen and Lebel’s (2012) contention is generally justifiable, a well-established area within which comparisons between subjective scores (self-reported) and objective scores (task-based) can be made is intelligence. To our knowledge, such research has not yet been conducted at the true score level, in conjunction with a multi-dimensional measure of SDR. Consequently, the primary purpose of this investigation was to estimate the true score associations between self-reported intelligence, task-based intelligence, and SDR. An observed increase in the association between self-reported intelligence and task-based intelligence, controlling for the effects of SDR on self-reported intelligence, was considered supportive of the contention that self-report measures may be genuinely influenced by SDR, and that the measurement of SDR may possess validity as an indicator of socially desirable responding, rather than represent substantive trait variance.
Socially Desirable Responding: Problem or Not?

Socially desirable responding (SDR) is the tendency to modulate responses to questions in order to look good and/or avoid looking bad (Edwards, 1957). Given that most self-report measures in psychology tend to be associated with a substantial amount of face validity (Burger, 2008), it is reasonable to expect that respondents would be able to discern the scoring direction of an item and respond in a manner to suit their goals (Cattell & Warburton, 1967). In both low-stakes and high-stakes psychological testing, long-standing concerns have been raised about the possibility that self-reported test scores may be contaminated by SDR (Paulhus, 1991). However, whether SDR is in fact a problem for self-report measurement remains a contested issue (Tracey, 2016).

A substantial amount of empirical research supports the notion that many self-report measures are meaningfully susceptible to SDR, at least when examined experimentally. Most experimental studies have employed a between-subjects design, such that they compare the self-report scores of job applicants (or participants instructed to try to get a job) against incumbents (or a control group). Based on a meta-analysis of Big Five dimensions, Birkeland, Manson, Kisamore, Brannick, and Smith (2006) reported that job applicants score as much as approximately half of a standard deviation more attractively than incumbents (e.g., Conscientiousness $d = .45$). Correspondingly, Alliger and Dwight (2000) reported a meta-analytic effect size closer to a full standard deviation for integrity tests. At least superficially, the experimental research should be regarded as fairly incontrovertible evidence that SDR is a problem for the measurement of dimensions via self-report, particularly those items associated with an appreciable degree of perceived social value. However, several arguments have been articulated to negate the implications of the experimental work,
including the possibility that all testees engage in approximately the same amount of SDR, thus preserving the rank-order of test scores (Lautenschlager, 1994). However, perhaps the most compelling counter-argument is based on the results, or lack thereof, derived from the correlational SDR research.

The majority of the correlation research in the area is predicated upon the contention that a self-report measure’s criterion-related validity should be enhanced by the inclusion of a measure of SDR to the model (e.g., personality as a predictor of job performance). In statistical terms, such an effect is known as suppression (Horst, 1941). Based on an extensive review of the personality literature, Ones, Viswesvaran, and Reiss (1996) created a meta-analytically derived true score correlation matrix between the Big Five dimensions of personality, SDR, and job performance. Ones et al. (1996) failed to find evidence to suggest that SDR operated as a suppressor with respect to the criterion-related validity for any of the Big Five dimensions. For example, the following zero-order true score correlations were reported by Ones et al. (1996):
Conscientiousness and job performance, $r = .23$; Conscientiousness and SDR, $r = .20$; and SDR and job performance, $r = .01$. The corresponding semi-partial correlation between Conscientiousness and job performance, controlling for the effects SDR on Conscientiousness, was reported at $r = .23$, i.e., unchanged from the zero-order correlation. As similar effects were reported for the other Big Five dimensions, Ones et al. (1996) concluded that “...attempts to control for social desirability are unwarranted” (p. 669).

More recently, McGrath et al. (2010) conducted a meta-analysis based on studies that evaluated directly the response bias hypothesis via correlational techniques, rather than create a meta-analytically derived correlation matrix (as per Ones et al., 1996). McGrath et al. (2010) used the term ‘response bias hypothesis’ to
represent the notion that the predictive capacity of a substantive predictor should be enhanced by the inclusion of a valid indicator of bias to the analysis (a suppressor effect). Based on the analysis of 41 previously published studies, McGrath et al. (2010) concluded that clear evidence for the response bias hypothesis remained elusive, as there was little evidence to suggest consistent suppressor effects across a number of criterion-validity categories.

For example, based on the 10 studies that included self-reported personality ratings, spouse/informant personality ratings, and SDR, McGrath et al. (2010) found that only two of the 10 studies reported larger semi-partial correlations than corresponding zero-order correlations. McGrath et al. (2010) acknowledged the possibility that the absence of effects may have been the consequence of SDR measures with poor validity. It is noteworthy that many of the studies used the Marlowe-Crowne Scale (Crowne & Marlowe, 1960), or relatively under-researched validity scales unique to a particular inventory, as an indicator of SDR. However, one study included in the analysis reviewed above used a multi-dimensional measure of SDR – the Balanced Inventory of Desirable Responding (BIDR: Paulhus, 1991). Furthermore, the one study that used the BIDR reported larger semi-partial correlations, in comparison to zero-order correlations (Lee & Klein, 2002), which may regarded as supportive of the response bias hypothesis. Such an effect is consistent with the observation that the BIDR scale scores tend to be associated with greater validity as indicators of SDR, in comparison to other measures (e.g., Reid-Seiser & Fritzsche, 2001). Consequently, how SDR is measured may be considered an important consideration in the evaluation of the response bias hypothesis.

**Socially Desirable Responding: Measurement Issues**
Several measures of SDR have been published (Paulhus, 1991; Stöber, 2001; Vecchione, Alessandri & Barbaranelli, 2013). The Ones et al. (1996) and McGrath et al. (2010) meta-analyses were composed primarily of studies that used various versions of the Marlowe-Crowne Scale. The Marlowe-Crowne Scale has been evaluated critically on a number of grounds, including a relative lack of a theoretical underpinning (Paulhus, 1991), low internal consistency reliability (Beretvas, Meyers & Leite, 2002; Loo & Loewen, 2004), and an unclear factor structure (Barger, 2002; Leite & Beretvas, 2005). Additionally, the Marlowe-Crowne Scale was designed to measure a single dimension of SDR. By contrast, more sophisticated approaches to the measurement of SDR recognize its multi-dimensional nature (Bensch, Paulhus, Stankov, & Ziegler, 2017; Gignac, 2013). Consequently, researchers have been urged to consider the application of multidimensional measures of SDR in their research (Tracey, 2016).

A popular multidimensional approach to the measurement of SDR is the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991). The BIDR was designed to measure two relatively orthogonal dimensions of SDR: self-deceptive enhancement (SDE) and impression management (IM). The latent variable correlation between SDE and IM has been reported to range between $r \approx .15$ to $.40$ (Gignac, Karatamoglou, Wee, & Palacios, 2014; Paulhus, 1998), which suggests that total BIDR scores are insufficiently homogeneous to be interpreted. Paulhus (1984) found that the Marlowe-Crowne Scale related mostly to IM, although not exclusively so.

Theoretically, Paulhus and John (1998) proposed that SDE was consistent with an egoistic socially desirable responding mechanism. Paulhus (2002) described egoistic bias as the tendency to overestimate one’s abilities and to give the impression of being someone who can accomplish great achievements that are beyond the reach of others. From a values perspective, SDE is closely aligned with agency – a meta-concept
pertinent to the advancement of the self in social rankings (Paulhus & Trapnell, 2008). Correspondingly, the SDE subscale within the BIDR measures the degree to which people report the possession of unrealistic and/or socially valued abilities (e.g., ‘My first impressions of people usually turn out to be right’, ‘I never regret my decisions’; Paulhus, 1998). Empirically, SDE has been shown to correlate positively with overclaiming, narcissism and hindsight bias (Paulhus & Trapnell, 2008). Additionally, those who score high on SDE tend to claim abilities that are less likely to be corroborated by external raters (Paulhus, 1998).

Paulhus and John (1998) proposed that IM, in contrast to SDE, was more consistent with a moralistic bias socially desirable responding mechanism. Paulhus (2002) described moralistic bias as the tendency to deny impulses that are moderately deviant socially and to claim inordinately consistent altruistic behaviors. From a values perspective, IM is considered more closely aligned with communion - a meta-concept pertinent to the preservation of positive relationships (Paulhus & Trapnell, 2008). Correspondingly, the IM subscale within the BIDR measures the degree to which people claim very unlikely socially desirable behaviours that are typically related to social rule following (e.g., ‘When I hear people talking privately, I avoid listening’) or deny ever engaging in relatively minor socially disapproved behaviours (e.g., ‘I have said something about a friend behind his or her back’). Empirically, the IM subscale within the BIDR has been found to correlate positively and more substantially with the Agreeableness and Conscientiousness dimensions of personality, in comparison to SDE (Graziano & Tobin, 2002; Reid-Seiser & Fritzsche, 2001).

One of the intractable limitations associated with the conventional measurement of SDR is the possibility that SDR measures may represent substantive (personality) trait variance. For example, McCrae and Costa (1983) administered the NEO Inventory,
the Marlowe-Crowne Scale, and the Lie scale from Eysenck Personality Inventory (EPI) to a sample of 215 people from the general community. McCrae and Costa (1983) found that both self-reported and observer-reported scores from the SDR scales correlated with the self-reported personality traits roughly to the same degree. The results were interpreted to suggest that SDR measures, such as the Marlowe-Crowne Scale and the Lie scale from the EPI, measure a substantial amount of personality trait variance, an effect that has been replicated across several studies (Kurtz, Tarquini, & Iobst, 2008; Pauls & Stemmler, 2003). However, it is worth noting that the self-other correlations for conventionally conceived personality traits (Big Five) tend to be stronger than the self-other correlations for measures of SDR, as measured via the BIDR, which suggests that scores from more modern SDR measures may be, to some degree, distinct from personality measures (Lönnqvist, Paunonen, Tuulio-Henriksson, Lönnqvist, & Verkasalo, 2007). Ultimately, however, Paunonen and Lebel (2002) suggested that it was probably impossible to know the degree to which SDR scores represented response bias versus substantive trait variance. Although such a contention may be relevant for the area of personality, strictly defined, one area that would allow for the objective verification of self-report scores is intelligence, as well-validated, task-based measures of intelligence are readily available.

**Previous Empirical Research**

To our knowledge, only one investigation has measured self-reported intelligence (SRIQ), task-based intelligence (TBIQ), and SDR. Recently, Leising, Locke, Kurzius and Zimmermann (2016) administered a two-item measure of SRIQ, a small battery of task-based intelligence tests, and a measure of self-deceptive enhancement (not from the BIDR) to 201 undergraduate university students. They reported a positive correlation between SRIQ and TBIQ ($r = .26$), between SRIQ and SDE ($r = .26$), and a
non-significant correlation between SDE and TBIQ ($r = -0.04$). Leising et al. (2016) reported no meaningful increase in the association between SRIQ and TBIQ ($\beta = 0.27$ vs. zero-order $r = 0.26$), when they regressed SRIQ onto both TBIQ and SDE. Consequently, Leising et al. (2016) concluded that the response bias hypothesis was not supported.

However, the pattern of correlations reported by Leising et al. (2016) was suggestive of classical suppression (Paulhus et al., 2004), as the SDE variable had an association with SRIQ, but not with TBIQ. Consequently, the regression model employed by Leising et al. (2016) was arguably not optimal, in this case, as it implied the expectation of negative suppression (Paulhus et al., 2004). Arguably, the inclusion of SDE in a study relevant to the response bias hypothesis should be focused more simply upon the “purification” of the predictor variable. Stated alternatively, the issue at hand with respect to the influence of SDE on SRIQ is entirely measurement in nature, rather than consistent with the postulation of a mediation mechanism. Consequently, the semi-partial correlation approach to testing a suppressor effect hypothesis may be regarded as more appropriate, in this case (McGrath, 2010; Smith, Ager Jr & Williams, 1992; Velicer, 1978).

The semi-partial correlation approach to suppressor effect hypothesis testing specifies that the semi-partial correlation should be numerically larger than the zero-order correlation (Smith, Ager Jr & Williams, 1992; Velicer, 1978). Researchers tend not to test the difference between the zero-order and semi-partial correlation statistically, as the standard error of the difference between two such correlations has not yet been established. In practice, a simple numerical increase is considered evidence for suppression (Ludlow & Klein, 2014). However, based on simulation research, Paunonen and Lebel (2012) suggested that increases in $r^2$ of approximately 1% and 2% may be considered moderate and large suppressor effects, respectively.
In addition to the application of the more optimal semi-partial correlation approach to the testing measurement focused classical suppressor effect hypotheses, we note that the scores included in the multiple regression analysis reported by Leising et al. (2016) were far from perfectly reliable. Specifically, the following internal consistency reliabilities were reported: TBIQ = .60, SDE = .77, and SRIQ = .79. As internal consistency reliability is known to impact the results of regression analyses (Osborne & Waters, 2002), and the semi-partial correlation approach may be regarded as more appropriate, we re-tested the response bias hypothesis on the attenuated and disattenuated correlations (Nunnally & Bernstein, 1994) associated with the Leising et al. (2016) study.

As can be seen in Figure 1 (Panel A, Model 1), the (attenuated) semi-partial correlation between SRIQ and TBIQ, controlling for the effects of SDE on SRIQ, was \( r = .28 \) (\( r^2 = .078 \)), which was numerically larger than the zero-order correlation (Panel A, Model 2) between SRIQ and TBIQ, i.e., .26 (\( r^2 = .068 \)). In percentage of variance terms, the suppression effect amounted to a 1.0% increase in shared variance between SRIQ and TBIQ.

More substantial evidence for suppression was observed based on the disattenuated correlations. Specifically, as can be seen in Figure 1 (Model 1, Panel B), the disattenuated (corrected) semi-partial correlation between SRIQ and TBIQ, controlling for the effects of SDE on SRIQ, was \( r_c = .42 \) (\( r_c^2 = .176 \)), which was numerically larger than the disattenuated zero-order correlation (Panel B, Model 2) between SRIQ and TBIQ, i.e., \( r_c = .38 \) (\( r_c^2 = .144 \)). In percentage of variance terms, the suppression effect amounted to a 3.2% increase in shared true score variance between SRIQ and TBIQ. The pattern of results for both the attenuated and
disattenuated correlations was consistent with the observation of classical suppression (Paulhus et al., 2004; Paunonen & Lebel, 2012).

In addition to not taking internal consistency reliability into consideration, Leising et al. (2016) measured SDR with a novel, unidimensional scale designed to measure only SDE. Given the theoretical and empirical features that distinguish SDE from IM, as reviewed above, it is important to consider which type of response bias may impact different types of self-report measures. In the current investigation, SRIQ was the self-reported dimension of interest. Specifically, we were interested in both self-reported intellectual intelligence (SRIQ) and self-reported emotional intelligence (SREI). Consequently, as items within typical measures of SRIQ and SREI are clearly relevant to the possession or demonstration of abilities (agency), we hypothesized that both SRIQ and SREI would be impacted principally by SDE. By contrast, given that interpersonal dimensions (communion), such as emotional awareness of others and emotional management of others, are integral to many models and measures of EI (Neubauer & Freudenthaler, 2005), we hypothesized that IM would have an appreciable impact only on SREI, independently of the effects of SDE. In light of the above, it was considered useful to test the response bias hypothesis with a full, multidimensional measure of SDR, task-based measures of intelligence (IQ and EI), and validated measures of SRIQ and SREI within a true score data analytic framework (i.e., disattenuated for imperfect reliability).

Summary

There is a substantial amount of evidence to suggest that SDR may not be a problem for self-report measurement, at least when examined from the correlational perspective (Tracey, 2016). Furthermore, it has been contended that it is probably impossible to know whether a high SDR score represents response bias or a genuine
disposition (Paunonen & Lebel, 2012). To overcome the hurdles associated with discerning a person’s true personality, we tested the response bias hypothesis within a true score framework with self-report measures of intelligence (IQ and EI), task-based measures of intelligence (IQ and EI), and a relatively well-established, multi-dimensional measure of SDR (i.e., the BIDR). Finally, we focused upon the semi-partial correlation approach to response bias hypothesis testing. Based on the literature reviewed above, the following seven hypotheses were tested: (1) SRIQ and SREI will be correlated positively with SDE; (2) only SREI will be correlated positively with IM; (3) TBIQ and TBEI will be uncorrelated with SDE and IM; (4) SRIQ will be correlated positively with TBIQ; (5) SREI will be correlated positively with TBEI; (6) the association between SRIQ and TBIQ will increase, when controlling for the hypothesized effect of SDR (SDE) on SRIQ; and (7) the association between SREI and TBEI will increase, when controlling for the hypothesized effect of SDR (SDE and IM) on SREI.

**Method**

**Sample**

The sample consisted of 253 first-year undergraduate university students (64% female; mean age = 20.21; SD = 3.36). The participants were recruited from a first-year undergraduate psychology research pool within a large, English speaking university in Australia. The students participated in the research voluntarily for a small amount of extra course credit. Although information on ethnicity was not obtained from the participants, the university student body is known to be populated from a primarily white Anglo-Saxon background. One item value was found to be missing completely at random (Little’s MCAR Test, p > .05) and was replaced via maximum likelihood expectation maximization.

**Measures**
Given testing time constraints, it was not feasible to administer comprehensive inventories of all of the dimensions of interest. However, in order to help ensure measurement symmetry (Wittmann, 1988), tests of roughly comparable breadth were selected for the respective task-based and respective self-report measures.

**Intellectual intelligence (IQ).** Task-based intellectual intelligence was measured with a slight adaptation to the Similarities subtest from the WAIS-IV (Wechsler, 2008). The adaptation consisted of not following the instruction to prompt testees for more details for partially credited responses. The adaptation was implemented to save administration time. Internal consistency reliability was estimated at .66. Self-assessed intellectual intelligence was measured with the Self-Report Intelligence Questionnaire (SRIQ; Gignac, Stough, & Loukomitis, 2004). The SRIQ is a 9-item questionnaire (5-point Likert scale; strongly disagree to strongly agree). Three of the items are keyed negatively. Each item is roughly representative of one of the subtests within the Wechsler Adult Intelligence Scales (e.g., ‘I know a lot of worldly facts’; ‘I am poor at solving logical problems’). The ninth item is a general indicator ('I am intelligent'). Internal consistency reliability was estimated at .70.

**Emotional intelligence (EI).** Task-based emotional intelligence (TBEI) was measured with the brief version of the Situational Test of Emotional Understanding (STEU; Allen, Weissman, Hellwig, MacCann, & Roberts, 2014; MacCann & Roberts, 2008). The brief STEU consists of 19 vignettes designed to represent a situation that can be expected to elicit an emotion from the principal character in the vignette. The testee must select from five response alternatives the most appropriate emotion for the vignette. Based on the results of several investigations (Austin, 2010; Libbrecht & Lievens, 2012; Schlegel, Fontaine, & Scherer, 2017), the STEU may be considered a valid indicator of ability-based EI, on the basis of its positive correlation with the Mayer-
Salovey-Caruso Emotional Intelligence Test (MSCEIT), a social theory of mind ability test, and an emotional recognition test, for example. Based on our psychometric analysis, item 4, which has ‘angry’ as the correct response, was found to be associated with particularly poor psychometric properties. Consequently, we scored ‘frustrated’ rather than ‘angry’ as the correct response for item 4, as it yielded better psychometric properties and appeared somewhat more plausible, based on our reading of the vignette. Internal consistency reliability was estimated at .46.

*Self-assessed emotional intelligence* was measured with the 14-item short-form of the general version of the 70-item Genos Emotional Intelligence Inventory (Genos EI; Palmer, Stough, Harmer & Gignac, 2009). Although self-report in nature, Genos EI is arguably better characterized as a self-report measure of EI, rather than a mixed-model (combination of abilities, traits, self-esteem, well-being, for example) measure of EI (see Petrides, 2011), as the items are all closely aligned with the capacity to use EI skills (Gignac, 2010). Example items include, ‘I am effective in helping others feel positive,’ and ‘I fail to handle stressful situations effectively’. Six of the items are keyed negatively. Internal consistency reliability was estimated at .78.

**Socially desirability.** Socially desirable responding (SDR) was measured with the Balanced Inventory of Desirable Responding (BIDR), Version 7 (Paulhus, 1998). The BIDR consists of 40 items (half keyed negatively) measured on a 7-point Likert scale (1 = not true, 3 = somewhat, 7 = very true). In accordance with Paulhus (1991), after reverse scoring, the items were scored dichotomously such that responses of 6 or 7 were recoded into a value of 1 and all other responses were recoded into a value of 0. Half of the items measure self-deceptive enhancement (SDE) and the other half measure impression management (IM). In this sample, internal consistency reliability was estimated at .73 and .77 for the SDE and IM subscales, respectively.
**Procedure**

Data were collected as part of a larger study relevant to prospective memory. After reading the information sheet, participants signed the consent form. Next, participants completed the questionnaires via computer. Finally, the participants completed the cognitive ability testing with a test administrator. In addition to the Similarities subtest, the participants completed two tests of prospective memory and an author recognition test. All testing was completed individually and within approximately 45 to 50 minutes.

**Data Analysis**

As correlations are affected by measurement error (Nunnally & Bernstein, 1994), we employed a true score maximum likelihood approach to the estimation of the associations between the dimensions of interest. Specifically, each of the six test scores were modelled as single-indicator latent variables within Amos (Arbuckle, 2014). In accordance with Jöreskog and Sörbom (1982), the single-indicator latent variable uniqueness terms were fixed to $S^2*(1$-reliability), where $S^2$ was equal to each composite scores total variance and reliability was equal to internal consistency reliability of the respective tests’ scores (i.e., omega hierarchical via parcels; Gignac, 2014). The test score variances and reliabilities are reported in Table 1.

In order to evaluate the response bias hypothesis, classical suppression was tested via the semi-partial correlation approach (Smith, Ager Jr & Williams, 1992; Velicer, 1978). Specifically, first, we tested a simple, bivariate correlation model to estimate the zero-order true score correlation ($r_c$) between the self-reported dimension (i.e., SRIQ and SREI, separately) and the task-based dimension (TBIQ and TBEI, separately). Next, we estimated the true score semi-partial correlation (semi-partial $r_c$) between the respective self-reported and task-based dimensions, controlling for the
effects of SDR on the self-reported dimensions. Classical suppression was considered indicated, in the event that the squared semi-partial correlation was found to be larger than the corresponding squared zero-order correlation. Additionally, a well-fitting ($\chi^2 p > .05$) semi-partial correlation model was considered supportive of classical suppression, as it would imply that the association between SDR and the task-based measure of intelligence was not significantly different from zero. Because the null hypothesis is difficult to support within the frequentist hypothesis testing paradigm, Bayes factors were estimated for the relevant correlations with the ‘BayesFactor’ package (version 0.9.8; Morey, Rouder & Jamil, 2015) within R (R Core Team, 2016). Furthermore, as an inferential test of the difference between a squared zero-order correlation and a squared semi-partial correlation has not yet been established, we focused upon effect size. Specifically, we interpreted increases in $r^2$ of approximately 1% and 2%, respectively, as consistent with moderate and large suppressor effects (Paunonen & Lebel, 2012). Finally, as recommended by an anonymous reviewer, we also tested a competing ‘mediating suppression’ model (a.k.a., co-operative suppression; Cohen & Cohen, 1975), whereby the task-based measure of intelligence was specified to have a direct effect onto the self-report measure of intelligence, as well as an indirect effect via socially desirable responding. Mediating suppression was considered unsupported (and classical suppression suggested), in the event that the indirect effect was observed to be non-significant statistically. The standardized indirect effects were tested for statistical significance via bootstrapping (2,000 resamples; bias-corrected accelerated percentile method).

**Results**

No outliers were identified based on the inter-quartile range outlier labelling rule with a 3.0 multiplier (Hoaglin & Iglewicz, 1987). Furthermore, as can be seen in Table
1, the data were relatively normally distributed (skew < 1.0). Consequently, the data were considered sufficiently normal for parametric statistical analyses.

Prior to conducting the primary analyses, it was considered useful to demonstrate that the SRIQ and SREI scores were associated with similar descriptive statistics, in order to allow justifiable comparisons. The means associated with the SRIQ ($M = 3.54$) and SREI ($M = 3.58$) scales were not found to differ statistically, based on a paired-samples t-test, $t(252) = 1.01, p = .314, g = .06$. Furthermore, the SRIQ and SREI variances ($S^2 = .26$ and $S^2 = .20$, respectively) were not found to differ statistically significantly, based on a Levene’s test, $F(1, 504) = 3.32, p = .069$. Thus, for the sample on average, the participants were considered to have self-reported essentially equal levels of intellectual intelligence and emotional intelligence.

**Intelligence and SDR: Inter-Relations**

As can be seen in Table 1, SRIQ correlated statistically significantly with SDE ($r_c = .44, p < .001$), but not with IM ($r_c = .09, p = .306; BF_{01} = 2.75$), as hypothesized. Furthermore, SDE and IM were correlated positively, $r_c = .39, p < .001$. A test of the difference between the SDE and IM correlations with SRIQ was found to be statistically significant, $z = 5.39, p < .001$. Thus, SDE was found to be more substantially related to SRIQ than IM, as hypothesized. In order to estimate the unique effects of SDE and IM on SRIQ, SRIQ was regressed onto the SDE and IM dimensions, which yielded a disattenuated multiple $R^2_c = .201, p < .001, 95\% CI: .07/.33$. Thus, 20.1% of the true score variance in SRIQ was accounted for by the two SDR dimensions. However, as hypothesized (and expected based on the IM by SRIQ zero-order correlation), only SDE was found to be a statistically significant contributor to the model: SDE, $B_c = .08, \beta_c = .48, p < .001$, semi-partia\(l \ r_c = .43$; IM, $B_c = -.01, \beta_c = -.10, p = .310$, semi-partia\(l \ r_c = -.10$. 
In contrast to SRIQ, SREI correlated significantly with both SDE ($r_c = .58$, $p < .001$) and IM ($r_c = .29$, $p < .001$), as hypothesized. A test of the difference between the SDE and IM correlations with SRIQ was found to be statistically significant, $z = 2.96$, $p = .003$. Thus, SDE correlated more substantially with SREI than IM. Next, SREI was regressed onto the SDE and IM dimensions, which yielded a disattenuated multiple $R_c^2 = .346$, $p < .001$, 95%CI: .19/.50. Thus, 34.6% of the true score variance in SREI was accounted for by the two SDR dimensions. However, as per SRIQ, only SDE was found to be a statistically significant contributor to the model: SDE, $B_c = .09$, $\beta_c = .56$, $p < .001$, semi-partial $r_c = .46$; IM, $B_c = .01$, $\beta_c = .08$, $p = .366$, semi-partial $r_c = .08$.

In contrast to SRIQ and SREI, neither SDE nor IM correlated statistically significantly with task-based IQ or task-based EI (see Table 1), as hypothesized. The following null/alternative Bayes factors were estimated: TBIQ*SDE, $BF_{01} = 5.39$; TBIQ*IM, $BF_{01} = 6.00$; TBEI*SDE, $BF_{01} = 1.70$; TBEI*IM = 2.19. Thus, with respect to the TBIQ, the data supported the null hypothesis over an effect model by a Bayes factor value of greater than 5-to-1. By comparison, with respect to TBEI, the evidence in favor of the null was weaker, although the null hypothesis was nonetheless approximately two times more likely than an effect model.

**Convergent Validity: Multi-Measurement**

The true score correlation between SRIQ and TBIQ was positive in direction but not significant statistically, $r_c = .16$, $r^2_c = .026$, $p = .082$. By contrast, the true score correlation between SREI and TBEI was both positive and statistically significant, $r_c = .35$, $r^2_c = .122$, $p < .001$. Next, we tested the suppressor effect hypotheses via the semi-partial correlation method.

**Suppression**
The true score semi-partial correlation between SRIQ and TBIQ, controlling for the effects of SDE on SRIQ, was positive and statistically significant, \( r_c = .19, r_c^2 = .036 \) \( p = .045 \). Additionally, the model was associated with excellent fit, \( \chi^2(1) = 0.31, p = .579, \) CFI = 1.00, RMSEA = .000. Furthermore, the percentage of true score variance shared by SRIQ and TBIQ, controlling for the effects of SDE, was numerically larger than the corresponding SRIQ and TBIQ zero-order squared correlation, \( \Delta r_c^2 = .010 \), which was considered a moderate sized suppressor effect. Finally, the competing mediating suppression model yielded a non-significant standardized indirect effect \( \beta = -.02, p = .542, 95\%CI: -.12/.05 \). Thus, again, the classical suppressor effect hypothesis with respect to IQ was considered supported.

Next, the true score semi-partial correlation between SREI and TBEI, controlling for the effects of SDE on SREI, was positive and statistically significant, \( r_c = .38, r_c^2 = .144 \) \( p = .001 \). Additionally, the model was associated with excellent fit, \( \chi^2(1) = 1.09, p = .297, \) CFI = .999, RMSEA = .019. Furthermore, the percentage of true score variance shared by SREI and TBEI, controlling for the effects of SDE, was numerically larger than the corresponding SREI and TBEI zero-order squared correlation, \( \Delta r_c^2 = .021 \), which was considered a large sized suppressor effect. Finally, the competing mediating suppression model yielded a non-significant standardized indirect effect \( \beta = .06, p = .294, 95\%CI: -.06/.17 \). Thus, again, the classical suppressor effect hypothesis with respect to EI was considered supported.

**Discussion**

We reported the following key results in this investigation. SRIQ and SREI correlated positively with TBIQ and TBEI, respectively. However, the association between SRIQ and TBIQ was observed to be statistically significant, only once the influence of SDE was removed from SRIQ. Furthermore, the response bias hypothesis
was supported via the semi-partial correlation approach for both SRIQ (1.0% greater variance accounted for in TBIQ) and SREI (2.1% greater variance accounted for in TBEI) from the perspective of effect size.

**Intelligence and SDR**

The hypothesis that SRIQ would be correlated positively with SDR was supported in this investigation. Furthermore, only SDE, and not IM, was observed to yield a statistically significant zero-order association with SRIQ (i.e., $r = .44$), as hypothesized. Such a result is consistent with the finding that narcissism, a dimension similar to SDE, is correlated positively with the degree of overestimation of one’s intelligence (Gabriel, Critelli & Ee, 1994). The disattenuated correlation we calculated between SDE and SRIQ with the Leising et al. (2016) data (i.e., $r = .38$) is comparable in magnitude to the disattenuated effect found in this investigation. However, the non-significant (or negligible) association between IM and SRIQ remains to be confirmed, as Leising et al. (2016) did not administer the IM subscale.

The pattern of zero-order correlations between SDR (SDE and IM) and SRIQ observed in this investigation are consistent, theoretically, with the notion that SDE is driven by agency (Paulhus & John, 1998). By contrast, the motivational drives of communion (impression management) do not appear to be relevant to the response bias of SRIQ. It should be acknowledged, however, that the data in this investigation were collected in a relatively benign, low-stakes testing circumstance. There is some research to suggest that the association between IM and SRIQ may be moderated by the degree of perceived stereotype threat (von Hippel, et al. 2005). Consequently, the extent to which the results reported in this investigation would generalize to high-stakes testing circumstances remains to be determined.
In contrast to SRIQ, SREI evidenced statistically significant, positive, zero-order correlations with both SDE and IM, which suggests that responses to SREI items are driven by both agency and communion. Such an observation is consistent with our hypothesis that the perception of EI incorporates elements of both ability (agency) and the desire to affiliate with others socially (communion). However, when examined via multiple regression, IM failed to contribute uniquely to the model predicting SREI. Such a result suggests that the total effect of IM on SREI was mediated by SDE. However, it is difficult to substantiate a theoretically plausible model to represent such a mediated effect (IM → SDE → SREI). Instead, it would seem more plausible to suggest that there was a general SDR process that accounted for the total effect between IM and SREI. The inclusion of a third indicator of SDR could potentially evaluate these competing hypotheses, as a general SDR latent variable would require a minimum of three indicators.

Finally, although not hypothesized specifically, SREI was found to be impacted by SDR more substantially than SRIQ (Model $R_c^2 = .346$ vs. Model $R_c^2 = .201$; $p = .032$). The larger effect of SDR on SREI cannot be accounted for entirely by the shared variance between IM and SREI, as the association between SDE and SREI was statistically significantly larger than the association between SDE and SRIQ ($r_c = .44$ vs. $r_c = .58$, $p = .003$). Thus, it appears that agentic exaggerations are also more pronounced for emotional intelligence than for intellectual intelligence. Such an effect suggests that EI may be perceived as an even more desirable trait to possess than IQ, even though intellectual intelligence has been ranked consistently as one of the most attractive traits in a mate, behind only kindness (Buss et al., 1990; Goodwin & Tinker, 2002). However, we failed to observe a statistically significant difference between the SRIQ and SREI means (averaged across items). Admittedly, the SRIQ and SREI 5-point
Likert scales were not associated with the same verbal anchors, therefore, fully valid comparisons cannot be made. Consequently, it would be useful to evaluate the rated social desirability (and attractiveness) of intellectual versus emotional intelligence in a purpose designed future investigation.

**Convergent Validity: Multi-Measurement**

The true score correlation between SRIQ and TBIQ was estimated at .16 in this investigation. Although not significant statistically (95%CI: -.04/.33), the correlation was within sampling variability of the meta-analytically estimated true score effect of .30 (Freund & Kasten, 2012). Thus, it is almost undoubtedly the case that SRIQ and TBIQ are not measuring entirely the same constructs, as the correlation is too small to suggest so. Furthermore, the possibility of approximating a person’s objective, maximal, intellectual intelligence via self-report is untenable, at least with current measures (Paulhus, Lysy & Yik, 1998). It would be interesting to determine whether SRIQ is associated with any unique predictive validity, independently of the effects of TBIQ.

In contrast to the intellectual intelligence data, the true score correlation between SREI and TBEI was statistically significant (.35, *p* < .001). Such an effect is essentially consistent with the effect sizes reported in previous studies, although they were not disattenuated for imperfect reliability (*r* = .20 to .30; Brackett & Mayer, 2003; Brannick, Wahi, Arce, Johnson, Nazian, & Goldin, 2009). In contrast to previous investigations, this is the first to use Genos EI and the Situational Test of Emotional Understanding to estimate the convergent validity between SREI and TBEI. Thus, the moderate amount of shared variance between SREI and TBEI appears to be generalizable. Given the popularity of EI research, it is surprising that more combined SREI and TBEI studies have not been conducted. Researchers appear to endorse one
mode of measurement or the other, not both (Mayer, Salovey & Caruso, 2008). It remains to be determined whether the predictive validity associated with self-reported EI is mediated by ability-EI (and/or vice versa). A psychometric meta-analysis (Hunter & Schmidt, 2004) of the association between SREI and TBEI would be useful. It should be emphasized, however, that important theoretical distinctions between trait-EI and ability-EI have been widely accepted in the literature (Petrides, 2011). Thus, it should not be expected that the two types of measures will evidence a substantial degree of convergence.

**Response Bias Hypothesis**

The semi-partial correlation approach to response bias hypothesis testing suggested that the effects between self-reported intelligence and task-based intelligence were suppressed due to SDR by between 1% and 2.1%, for IQ and EI, respectively. Consequently, from an effect size perspective (see Paunonen & Lebel, 2012), the response bias hypothesis was supported in this investigation. Ideally, a dedicated statistical analysis of the difference between a squared zero-order correlation and a squared semi-partial correlation would be available to test a suppressor effect hypothesis via the semi-partial correlation approach. Smith et al. (1992) noted the difficulties with developing such a statistical test. Presumably, a valid statistical test may be derivable from the delta method (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Future research in this area is encouraged.

The fact that SDR correlated near zero or weakly with the task-based measures of intelligence, as hypothesized, is a critical observation in the context of a classical suppression effect (Horst, 1941; Paulhus et al., 2004). Weak correlations between SDR and objective performance are expected, because people who score high on valid measures of SDR are not considered to possess genuinely the claimed traits and/or
abilities (Paulhus, 1998; Edwards, 1957). Consequently, in the context of the response bias hypothesis, so long as a SDR measure is found to correlate positively with the criterion of interest (e.g., job performance), the validity coefficients will necessarily be reduced, which is non-supportive of the response bias hypothesis (see Vecchione, et al. 2013, for an example with conscientiousness and job performance). The results of this investigation underscore the importance of measuring SDR from a multi-dimensional perspective, as IM was not found to be a unique predictor of SRIQ or SREI.

Correspondingly, Paulhus (1984) contended that the Marlowe-Crowne Scale will likely underestimate the effects of SDR with criteria, as it is mostly related to only the IM spectrum of SDR.

**BIDR: Substance, Style, or Both?**

The results of the suppressor effect analyses suggest that scores from the BIDR may be regarded as at least partially valid representations of individual differences in socially desirable responding. Much of the evidence supportive of the notion that SDR measures are indicators of genuine trait variance is based on the similarities between with self-report and other-report correlations with criteria (e.g., McCrae & Costa, 1983). It is noteworthy, however, that Lönnqvist et al. (2007) reported self-spouse correlations of .35 and .33 for the SDE and IM scales from the BIDR (scored polytomously). By contrast, the reported mean self-spouse correlation was .52 for the Big 5 dimensions (Lönnqvist et al., 2007). Thus, although SDR may be imbued by some substantive trait variance, the scale scores are not entirely akin to typical personality trait scores. All things considered, scores from the BIDR (scored dichotomously) may be considered both style and substance, as acknowledged by others (Holden & Passey, 2010). Further developments and refinements in the measurement of SDR may help improve the ratio of style to substance.
Implications: Is SDR a Problem?

The small, absolute increases in criterion-related validity reported in this investigation are easy to dismiss as inconsequential. In part, it is on this basis that the threat of SDR for self-reported measures has been dismissed (e.g., Leising et al., 2016). However, the substantial amount of experimental research supportive of the response bias hypothesis (see Holden & Book, 2012) is arguably incompatible with the correlational response bias research. Consequently, some have argued that the impact of response bias on a criterion-related validity coefficient may not be an appropriate representation of the magnitude of the problem (Alliger & Dwight, 2000). However, rather than generate a novel, quantitative index, it is possible that we simply need to reframe our expectations and interpretations of response bias and the existing validity coefficient.

Notably, the simulation research conducted by Paunonen and Lebel (2012) demonstrated that a moderate amount of SDR contamination in a predictor variable depresses a true (non-SDR contaminated) validity coefficient from .20 to .18. By comparison, a large amount of SDR contamination in a predictor variable depresses a true validity coefficient from .20 to .17. Paunonen and Lebel (2012) found that larger validity coefficient depressions were observed for larger true validity coefficients, when manipulated by the same two levels of SDR contamination (moderate and large). Thus, the effect of SDR in correlation research, to-date, may have been partly obscured by the fact that many of the true validity coefficients researchers attempt to identify are relatively small, at least with existing measures. Additionally, small suppressor effects in absolute terms may have been disregarded, unjustifiably.

---

¹ The effects were proportionally the same, however, in percentage of variance terms.
In addition to the simulation research, Rosse, Stecher, Miller, and Levin (1998) found that correlations in the order of .30 to .40 between SDR and self-reported personality can lead to non-negligible differences in rank-ordering, which could be reasonably expected to impact hiring decisions. More recently, Donovan, Dwight, and Schneider (2014) used a within-subjects design to evaluate the degree of SDR engaged by job applicants. Most noteworthy, Donovan et al. (2014) found that those job applicants who were ultimately hired self-reported less trait relevant goal motivation four months later during a training seminar (low-stakes). Furthermore, those who were hired that exhibited substantial levels of SDR pre-employment performed less well in the job. Thus, in light of the above, there is arguably compelling evidence to suggest that self-report measures are susceptible to SDR to a non-negligible degree, even if the suppressor effects are small in absolute terms, as reported in this investigation.

In light of the simulation and applied research, increases in criterion-related validity equal to $r \approx .10$ ($r^2 \approx .01$) should probably be regarded as at least moderate in practical terms. Increases in validity equal to $r \approx .15$ ($r^2 \approx .02$) may be regarded large in practical terms. Thus, the suppressor effect of SDR on the convergent validity of SRIQ found in this investigation may be regarded as moderate. Furthermore, the suppressor effect of SDR on SREI may be regarded as large. From an applied perspective, the implications of these results suggest that the use of SREI measures in applied contexts (e.g., personnel selection), without a correction for SDR, may not be recommendable, as individual differences in SDR will likely lead to some poor decisions. As practitioners do not appear to use SRIQ measures as an indicator of objective intellectual intelligence, the practical implications for the SDR suppressor effects relevant to self-assessed intellectual intelligence are limited. It is also acknowledged that some may not view the absence of suppressor effects in previous research as
inconsistent with the experimental fake-good research on a number of bases, including the possibility that self-report dissimulation is not substantial in applied settings and/or that there are little individual differences in dissimulation in high-stakes assessment settings (e.g., Lautenschlager, 1994; Sackett, 2011).

**Limitations**

Although the sample size used in this investigation was respectable \((N = 253)\), it was not large. The sample sizes required to detect moderate and small suppressor effects with 80% power are 270 and 614, respectively. As Paunonen and Lebel (2012) demonstrated via simulation, only relatively small suppressor effect sizes can be expected (1% to 2%), even when there is a substantial amount of SDR bias within the data. Others have noted power concerns associated with the correlational approach to the evaluation of the response bias hypothesis (Holden, 2008). Consequently, it would be beneficial to replicate the results reported in this investigation with another sample, preferably a relatively large one \((N > 500)\).

The university sample employed in this investigation was likely associated with some range restriction in the intelligence test scores (not possible to estimate validly, given the nature of IQ tests). Therefore, it is plausible to suggest that the degree of suppression reported in this investigation is an underestimation of the effect in the population. Consequently, a replication study with participants from the general community would be valuable.

Ideally, all measures administered in this investigation would have yielded respectably reliable scores \((> .70; \text{Nunnally & Bernstein, 1994})\). However, in this investigation, the brief Situational Test of Emotional Understanding yielded a test score reliability of .46, which implies that only 46% of the test’s variance was reliable variance. The STEU is not the only task-based subtest of EI to yield low internal
consistency reliability (i.e., < .70): it appears to be the norm, rather than the exception (Maul, 2011; Palmer et al., 2005; Śmieja, Orzechowski, & Stolarski, 2014). Work to improve the measurement of task-based EI is encourage.

Finally, task-based IQ and task-based EI were each measured with only one subtest. It is possible that a more comprehensive task-based measurement of intelligence would have revealed statistically significant correlations between SDR and TBIQ and TBEI, which would have reduced the amount of SDR suppression, all other things equal. However, based on a sample of 102 adults from the general community, Gudjonsson & Young (2011) reported non-significant associations between the two subscales of the BIDR and the WASI. Nonetheless, it would be useful to extend the results of this investigation with more comprehensive measures of intelligence. Conversely, improvements in the measurement of SDR may yield more substantially sized suppressor effects, if scores from the BIDR are demonstrated to be less than perfectly valid representations of SDR (almost undoubtedly the case).

**Conclusion**

Points of view about the impact of SDR on the self-reported measurement of psychological attributes have varied widely for many years, from red herring to serious problem. An arguably justifiable contemporary view, based on the more recent correlational research, and more modern, multi-dimensional measures of SDR, is that individual differences in the tendency to engage in SDR may be considered to impact the validity of self-reported scores, non-negligibly. Particularly if suppressor effects in the order of 1 to 2% are reframed as moderate to large in practical terms, despite their small, absolute magnitude.
References


Comparisons of trait and ability emotional intelligence measures in medical students. Medical education, 43, 1062-1068.


Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and


Horst, P. (1941). The role of predictor variables which are independent of the criterion. *Social Science Research Council, 48*, 431-436.


Lautenschlager, G.J. (1994). Accuracy and faking of background data. In G.S. Stokes,


Osborne, J.W. & Waters, E. (2002). Four assumptions of multiple regression that
researchers should always test. *Practical Assessment, Research and Evaluation*, 8(2).


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.


preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.


### Table 1

**Descriptive Statistics and Inter-Correlations**

<table>
<thead>
<tr>
<th></th>
<th>Inter-Correlations</th>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.</td>
<td>2.</td>
</tr>
<tr>
<td>1. SRIQ</td>
<td>--</td>
<td>.58</td>
</tr>
<tr>
<td>2. SREI</td>
<td>.43</td>
<td>--</td>
</tr>
<tr>
<td>3. TBIQ</td>
<td>.11</td>
<td>.09</td>
</tr>
<tr>
<td>4. TBEI</td>
<td>.14</td>
<td>.21</td>
</tr>
<tr>
<td>5. SDE</td>
<td>.31</td>
<td>.44</td>
</tr>
<tr>
<td>6. IM</td>
<td>.06</td>
<td>.22</td>
</tr>
</tbody>
</table>

*Note. N = 253; correlations below the main diagonal are observed (attenuated) based on ordinary least squares estimation; correlations above the main diagonal were disattenuated for imperfect reliability as estimated via maximum likelihood; correlations in bold were statistically significant (p < .05); SRIQ = self-report intelligence; SREI = self-report emotional intelligence; TBIQ = task-based intelligence (Similarities); TBEI = task-based emotional intelligence (Situation Emotional Intelligence Test); SDE = self-deceptive enhancement; IM = impression management; ω = coefficient omega; (1- ω)*S² = error variance.*
Figure 1

Re-analyses of Leising et al (2016); rectangles represent observed variables; circles represent latent variables; coefficients represent standardized effects; Panel A = effects attenuated due to imperfect reliability; Panel B = effects disattenuated for imperfect reliability; Model 1 = semi-partial correlation; Model 2 = bivariate correlation; classical suppression (and support for the response bias hypothesis) is considered consistent with an increase in $r^2$ from Model 2 to Model 1; SDR = socially desirable responding; SRIQ = self-reported intelligence; TBIQ = task-based intelligence.