


The Balanced Inventory of Desirable Responding Short Form (BIDR-16)

SAGE Open
 October–December 2015: 1–9
 © The Author(s) 2015
 DOI: 10.1177/2158244015621113
 sgo.sagepub.com


Claire M. Hart¹, Timothy D. Ritchie², Erica G. Hepper³,
 and Jochen E. Gebauer⁴

Abstract

Self-report studies often call for assessment of socially desirable responding. Many researchers use the Marlowe–Crowne Scale for its brief versions; however, this scale is outdated, and contemporary models of social desirability emphasize its multi-dimensional nature. The 40-item Balanced Inventory of Desirable Responding (BIDR) incorporates Self-Deceptive Enhancement (honest but overly positive responding) and Impression Management (bias toward pleasing others). However, its length limits its practicality. This article introduces the BIDR-16. In four studies, we shorten the BIDR from 40 items to 16 items, while retaining its two-factor structure, reliability, and validity. This short form will be invaluable to researchers wanting to assess social desirability when time is limited.

Keywords

Balanced Inventory of Desirable Responding (BIDR), impression management, Marlowe–Crowne, self-deceptive enhancement, social desirability

Socially desirable responding (SDR) in self-reports is a key concern for survey researchers. A critical assumption of self-report surveys is that respondents accurately bring to mind-relevant information and attempt to provide honest responses (McIntire & Miller, 2000). To the extent that individuals instead provide socially desirable responses (over-reporting positive behavior or under-reporting negative behavior), the validity of survey scores could be compromised.

Social desirability concerns are pervasive across disciplines where self-report questionnaires play an integral role in data collection. As such, attempts to control desirable responding have been made in psychology (Crowne & Marlowe, 1964), management and personnel selection (Goffin & Christiansen, 2003; Thompson & Phua, 2005), marketing (Steenkamp, de Jong, & Baumgartner, 2010), and medicine (Herbert, Clemow, Pbert, Ockene, & Ockene, 1995; Klesges et al., 2004).

Accordingly, social desirability scales are widely used to assess the extent to which individuals bias responses in a self-favoring manner and control for such distortions (Paulhus, 2002). Typically, such response biases are identified by administering an SDR scale alongside scales of interest. A non-significant association between SDR and focal scales implies the scale in question is free from response bias. In the case of a significant association, partialling out effects of SDR shows whether the scale predicts external criteria after variance attributed to social desirability is accounted for (Kam, 2013; Moorman & Podsakoff, 1992).

The ubiquitous problem of SDR has led to a scale proliferation (Paulhus, 1991; Uziel, 2010). The scale most commonly used in the past 60 years is the Marlowe–Crowne Social Desirability Scale (MCSDS; Beretvas, Meyers, & Leite, 2002; Crowne & Marlowe, 1964). This scale comprises 33 items that are either socially desirable but uncommon (approved behaviors) or socially undesirable but common (disapproved behaviors). Individuals scoring high on approved and low on disapproved behaviors display high SDR. According to Crowne and Marlowe (1964), SDR on their scale represents a single latent construct—need for approval.

Despite being the most frequently cited SDR scale, the MCSDS's length, outdated wording, and low reliability limits its practicality (Ballard, Crino, & Rubenfeld, 1988; Beretvas et al., 2002; Stöber, 2001). Several 10- to 20-item short forms of the MCSDS have been developed to address the first of these concerns (Ballard, 1992; Reynolds, 1982;

¹University of Southampton, UK

²Saint Xavier University, Chicago, IL, USA

³University of Surrey, Guildford, UK

⁴University of Mannheim, Germany

Corresponding Author:

Claire M. Hart, Centre for Research on Self and Identity, School of Psychology, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK.

Email: c.m.hart@soton.ac.uk



Stöber, 2001; Strahan & Gerbasi, 1972). However, with the exception of Stöber's version, item wordings are unchanged and, thus, remain outdated. Moreover, the MCSDS, including its short forms, represent SDR as unidimensional; item selection was based on items' correlations with the principal factor, need for approval. However, there is no clear evidence supporting the fit of a one-factor "need for approval" model to scores on the various MCSDS forms (Barger, 2002; Leite & Beretvas, 2005; Paulhus, 1984; Stöber, Dette, & Musch, 2002).

This issue of dimensionality in SDR has long been disputed (Crowne & Marlowe, 1964; Damarin & Messick, 1965; Sackeim & Gur, 1978). Paulhus (1984) conducted factor analyses of various SDR measures and consistently obtained two factors: impression management (IM) and self-deceptive enhancement (SDE). In Paulhus's (1984) original conceptualization, IM, similar to Marlowe and Crowne's need for approval, signifies a tendency to give inflated self-descriptions to an audience: a conscious dissimulation of responses to create a socially desirable image. Conversely, SDE represents a tendency to give honest but positively biased reports (Paulhus, 1984): a non-conscious inclination to perceive oneself favorably. Since his original proposal for a two-factor structure, Paulhus and Reid (1991) have argued for a three-factor structure, in which SDE can be further divided into self-deceptive enhancement and self-deceptive denial, and more recently, a two-tiered construct model crossing content (agentic vs. communal) with responsiveness to audience manipulation (public vs. private; Paulhus & Trapnell, 2009). The issue of dimensionality of SDR is ongoing (Gignac, 2013; Lanyon & Carle, 2007; Leite & Beretvas, 2005).

To operationalize the two-factor SDR model, Paulhus (1991, 1998) developed the Balanced Inventory of Desirable Responding (BIDR). The BIDR contains 40 items: 20 IM items and 20 SDE items. The BIDR is a robust measure showing satisfactory internal consistency and test-retest reliability (Paulhus, 1994). Many researchers who use the BIDR continue to calculate the two originally conceived subscales. These subscales have discriminant validity, with IM (but not SDE) showing sensitivity to variations in anonymity (higher scores in public than private), and SDE (but not IM) predicting overconfidence, hindsight, and overclaiming (Paulhus, 1994).

Paulhus (1984) highlights the implications of ignoring SDR's multidimensionality, arguing the absence of a correlation between a focal measure and a unidimensional SDR scale (e.g., MCSDS) does not necessarily mean there is no SDR in that measure. Efforts to control SDR must address both dimensions. The BIDR affords flexibility by allowing control of either one or both components, depending on the focal scale(s) of interest.

The BIDR enjoys widespread use across varied disciplines. However, many researchers may be reluctant to use the 40-item scale. The addition of a long scale to an existing

study may increase transient measurement errors, as respondents become frustrated or respond carelessly due to boredom or fatigue (Schmidt, Le, & Iles, 2003). Instead, researchers may opt to use a short form MCSDS which only captures the IM dimension of SDR. As far as we are aware, there are no English-language short forms of the BIDR (Leite & Beretvas, 2005). The validation of SDR scale short forms is a serious concern given the costs of including a long MCSDS or BIDR together with focal measures. This lack of valid yet practical scales may prevent researchers from identifying and controlling for unwanted SDR variance.

In this article, we report four studies in which we shorten the BIDR from 40 items to 16 items, retaining its two-factor structure, reliability, and validity. We hope the BIDR-16 can be implemented when a longer measure is impractical. Study 1 uses confirmatory factor analysis (CFA) on datasets containing the BIDR-40 to shorten the BIDR and provides preliminary construct validity evidence by showing comparable relationships between both BIDR forms and external correlates. Study 2 replicates the CFA findings using an independent sample with administration of the BIDR-16 only. Study 3 examines test-retest reliability of the BIDR-16. Study 4 cross-validates the BIDR-16 with external correlates.

We used the following external criteria: (a) self-enhancement measures to validate the short SDE scale, (b) a short form MCSDS to validate the short IM scale, and (c) the Big Five personality traits (John & Srivastava, 1999) to show divergent relations with SDE and IM. We expected the BIDR-16 would show correlational patterns consistent with the BIDR-40 in direction and magnitude. Concerning (a), overclaiming and overconfidence correlate positively with SDE but not IM (Paulhus, 1994). Such propensities are also characteristic of individuals high in subclinical narcissism and self-esteem (Campbell, Goodie, & Foster, 2004; Greenberger, Chen, Dmitrieva, & Farrugia, 2003). Accordingly, we anticipated (in Studies 1 and 4) that correlations of SDE with measures of self-esteem and narcissism would be positive and stronger than with IM. Concerning (b), unidimensional SDR measures typically correlate strongly with IM and weakly with SDE (Paulhus & Reid, 1991). In Study 4, we expected a short form MCSDS to correlate more strongly with IM than SDE using the BIDR-16. Concerning (c), SDE and IM show different relations with key personality traits. In a meta-analysis, Li and Bagger (2006) report that SDE correlated most strongly with emotional stability, followed by conscientiousness and extraversion, then agreeableness and openness. Conversely, IM correlated most strongly with conscientiousness and agreeableness, followed by emotional stability, extraversion, and openness. We expected to replicate these patterns with the BIDR-16 in Study 4.

Study 1

Study 1 examined datasets that included the BIDR-40 (Paulhus, 1991, 1998). We aimed to evaluate the BIDR-40

on model fit and dimensionality and to refine it into a shortened theory-grounded model consistent with the original 40-item version. A secondary aim was to provide preliminary validation for the BIDR-16 subscales by demonstrating equivalent relationships between the BIDR-40 and BIDR-16 and self-enhancement measures (self-esteem, narcissism).

Method

Participants. Eight datasets¹ contained 1,948 participants (1,479 women, four undisclosed; $M_{\text{age}} = 23.28$, $SD = 8.30$, range = 16–73). Of these, 854 were from the United Kingdom, 815 from the United States, and 279 from other countries, including Australia, Canada, Europe, and East Asia (one undisclosed).

Materials and procedure. Participants completed the BIDR-40 (Paulhus, 1991, 1998)² comprising 20 SDE items ($\alpha = .70$) and 20 IM items ($\alpha = .78$) rated on 7-point scales (1 = *strongly disagree*, 7 = *strongly agree*), following Stöber et al.'s (2002) and Kam's (2013) recommendations for continuous rather than dichotomous scoring.

Participants completed one of three self-esteem measures. Participants completing the Rosenberg (1965) Self-Esteem Scale (RSES; $\alpha = .88-.90$, $n = 1,299$) rated 10-items such as “I have a number of good qualities” (1 = *strongly disagree*, 4 [or 11] = *strongly agree*). Those completing the Single-Item Self-Esteem measure ($n = 212$; Robins, Hendin, & Trzesniewski, 2001) rated the item “I have high self-esteem” (1 = *disagree strongly*; 11 = *agree strongly*). Those completing the Self-Liking Self-Competence Scale-Revised ($\alpha = .91$, $n = 198$; Tafarodi & Swann, 2001) rated 16-items such as “I feel great about who I am” (1 = *disagree strongly*, 7 = *agree strongly*). We used standardized scores from these scales to compute a self-esteem index.

Participants completed one of two versions of the Narcissistic Personality Inventory (NPI; Raskin & Terry, 1988): the NPI-40 ($\alpha s = .80-.87$, $n = 564$), or NPI-15 (Schütz, Marcus, & Sellin, 2004; $\alpha s = .82$, $n = 760$). In each measure, participants choose which statement is most true of them (e.g., “I like to be the center of attention” [high-narcissistic]/“I prefer to blend in with the crowd” [low-narcissistic]). We used standardized scores from these scales to compute a narcissism index.

Analytic Strategy

Because data were collected with BIDR versions 6 and 7, we removed items that differed between versions. Using the remaining 36-items, we followed three analytic strategies practiced in CFA (Jöreskog, 1993) to identify an optimal subset of items. First, we used a *strictly confirmatory approach* to test the fit of a 36-item, two-factor model to the entire Study 1 sample and four subsamples (divided by nation and gender). Second, we used a *model generating strategy* to

modify and test increasingly refined models by removing weakly loading items, preserving the two-factor structure, and retaining/improving model fit. Finally, we used the *alternative model strategy*, to test whether the proposed two-factor model fit better than a one-factor model, consistent with SDR theory (Paulhus, 1984, 2002). All structural analyses were aimed at developing and confirming a theory-grounded model that is consistent with the BIDR-40.

We assessed goodness of fit of each CFA model using maximum likelihood chi-square (χ^2), goodness-of-fit index (GFI) and comparative fit index (CFI), root mean square error approximation (RMSEA), and standardized root mean residual (SRMR). A good-fitting model is indicated by a non-significant chi-square test, GFI and CFI indices of at least .90 (Bentler & Bonett, 1980), and RMSEA and SRMR indices below .08 (Hu & Bentler, 1998). Several authors have noted that the model chi-square test, due to its sensitivity to sample size, is unacceptably conservative (e.g., Bentler & Bonett, 1980). Given our large sample sizes, it was unlikely we would obtain non-significant chi-square tests.

Results and Discussion

A priori measurement model. First, we subjected the initial 36-item, two-factor model to CFA using LISREL 8.80 (Jöreskog & Sörbom, 2006) to analyze the covariance matrix. Indicators suggested a reasonable fit to the data, $\chi^2(593, N = 1,850) = 3,627.04$, GFI = .89, CFI = .84, RMSEA = .058 (90% confidence interval [CI] = [.057, .060]), SRMR = .06. In general, the CFI shows an undesirable feature—namely, this fit index decreases with an increasing number of indicators per latent variable (Kenny & McCoach, 2003). Thus, it is not surprising that only the RMSEA and the SRMR evidence good fit in the present CFA (18 indicators per latent variable). In an attempt to refine the BIDR, we adopted a model generating strategy and tested a series of two-factor models with fewer items.

Model generating: Refining the initial factor structure. We refined the model by eliminating items with factor loadings and $R^2 < .30$ (Brockway, Carlson, Jones, & Bryant, 2002). Although we would have preferred to maintain high reliabilities ($\alpha \geq .70$), previous research has demonstrated the internal consistency of both SDE and IM is typically below or around .70 (Li & Bagger, 2007). Our results for the BIDR-40 replicate such findings (Table 1). Nonetheless, using the model generating approach, we examined a two-factor model with 10 items per factor. This model evinced improved fit from the 36-item version, $\chi^2(169, N = 1,850) = 1,417.00$, GFI = .92, CFI = .88, RMSEA = .067 (90% CI = [.065, .071]), SRMR = .06. We further refined the model by testing eight items per factor. This model also provided acceptable fit, $\chi^2(103, N = 1,850) = 913.92$, GFI = .93, CFI = .88, RMSEA = .067 (90% CI = [.066, .074]), SRMR = .05. Table 2 displays item loadings for the resulting BIDR-16.

Table 1. Descriptive Statistics and Alpha Reliabilities for Studies 1-4.

Study	Sample (n)	Scale	Minimum	Maximum	M	SD	Alpha
Study 1	Total (1,948)	SDE	1.38	7.00	3.79	0.86	.66
		IM	1.00	6.75	3.65	0.96	.72
	UK (854)	SDE	1.38	6.50	3.68	0.81	.63
		IM	1.13	6.63	3.55	0.90	.71
	USA (815)	SDE	1.50	7.00	3.87	0.88	.66
		IM	1.00	6.75	3.71	1.01	.74
	Women (1,479)	SDE	1.38	6.75	3.74	0.84	.65
		IM	1.00	6.75	3.67	0.97	.73
	Men (465)	SDE	1.50	7.00	3.96	0.90	.67
		IM	1.00	6.50	3.59	0.96	.70
Study 2	Total (670)	SDE	1.25	7.75	4.30	1.14	.69
		IM	1.00	7.88	4.50	1.24	.71
Study 3	Total (352)	SDE (Time 1)	1.38	8.00	4.59	1.07	.82
		SDE (Time 2)	1.38	8.00	4.77	1.11	.67
		IM (Time 1)	1.75	8.00	4.53	1.09	.67
		IM (Time 2)	2.00	8.00	4.59	1.05	.66
Study 4	Total (708)	SDE	1.25	8.00	4.49	1.04	.64
		IM	1.00	7.75	4.30	1.19	.73

Note. SDE = self-deceptive enhancement; IM = impression management.

Table 2. Factor Loadings Per Sample for the Two-Factor BIDR-16 Model in Study 1.

Factor	Item	Sample				
		All ^a	Women ^b	Men ^c	UK ^d	USA ^e
SDE	Item 4 “not always honest” (r)	.46	.45	.49	.38	.52
	Item 5 “know why like things”	.42	.43	.34	.37	.41
	Item 10 “hard to shut off a disturbing thought”(r)	.40	.37	.43	.32	.47
	Item 11 “never regret decisions”	.56	.56	.55	.53	.53
	Item 12 “can’t make up my mind” (r)	.44	.40	.51	.41	.48
	Item 15 “completely rational”	.46	.47	.41	.47	.39
	Item 17 “confident in judgements”	.52	.55	.39	.58	.41
	Item 18 “doubted ability as a lover” (r)	.33	.31	.36	.32	.41
IM	Item 21 “sometimes tell lies” (r)	.56	.56	.57	.59	.55
	Item 22 “never cover up mistakes”	.47	.46	.53	.44	.46
	Item 23 “taken advantage of someone”(r)	.48	.46	.58	.48	.53
	Item 25 “sometimes try to get even” (r)	.44	.44	.40	.39	.49
	Item 27 “said something bad about a friend” (r)	.57	.60	.49	.52	.61
	Item 28 “avoid listening”	.48	.53	.33	.47	.50
	Item 36 “never take things”	.38	.38	.39	.40	.38
	Item 40 “don’t gossip”	.53	.56	.43	.47	.59

Note. Item numbers correspond to the BIDR version 6. The loadings are from the completely standardized solution of a maximum likelihood CFA. Missing values were deleted using LISREL 8.80 listwise procedures. BIDR = Balanced Inventory of Desirable Responding; SDE = self-deceptive enhancement; IM = impression management; CFA = confirmatory factor analysis.

^an = 1,850.

^bn = 1,405.

^cn = 440.

^dn = 814.

^en = 778.

Strictly confirmatory: Testing subsample generalizability. We assessed how well the 16-item, two-factor model fit across four subsamples: per nation (the United Kingdom, the United

States) and per gender (Table 2). For each nation and gender, the refined model fit reasonably well; the United Kingdom: $\chi^2(103, N = 814) = 471.50$, GFI = .93, CFI = .84,

RMSEA = .069 (90% CI = [.063, .075]), SRMR = .06; the United States: $\chi^2(103, N = 778) = 429.30$, GFI = .93, CFI = .91, RMSEA = .069 (90% CI = [.063, .075]), SRMR = .05; women: $\chi^2(103, N = 1,405) = 672.17$, GFI = .92, CFI = .90, RMSEA = .067 (90% CI = [.063, .072]), SRMR = .05; men: $\chi^2(103, N = 440) = 331.95$, GFI = .92, CFI = .83, RMSEA = .070 (90% CI = [.062, .079]), SRMR = .07

Alternative model: Is the BIDR unidimensional? We first tested a 36-item, one-factor model. This evinced a poorer fit to the data than the two-factor model: $\chi^2(594, N = 1,850) = 4,799.26$, GFI = .83, CFI = .77, RMSEA = .074 (90% CI = [.073, .076]), SRMR = .07, with a difference of 1 *df*, the $\chi^2_{\text{Difference}} = 1172.22$. As the $\chi^2_{\text{Difference}} > 1,000$, a *p* value cannot be computed; however, Akaike information criterion (AIC) values of the one- and two-factor models (6,798.94 and 4,489.57, respectively) suggest the two-factor solution produces a better fit to the data. We next tested the 16-item, one-factor model. This also evinced a poorer fit to the data than the two-factor model: $\chi^2(104, N = 1,850) = 1,571.27$, GFI = .88, CFI = .79, RMSEA = .099 (90% CI = [.096, .103]), SRMR = .07, with a difference of 1 *df*, the $\chi^2_{\text{Difference}} = 657.35$, *p* < .001. The one-factor model fit the data poorly for all subgroups.³

The two-factor BIDR-16 fit the data relatively well. Reduction of more than half the items did not weaken the structural validity of the BIDR; in fact, it reinforced it—AIC values for the short version were lower than for the full version (1,104.39 and 4,489.57, respectively).

External correlates. We garnered preliminary validity evidence for the BIDR-16 subscales (Table 3). Replicating past findings (Greenberger et al., 2003), we obtained a positive correlation between SDE and IM. SDE correlated more strongly with self-esteem than did IM, $z = 15.90$, *p* < .001.⁴ SDE also correlated more strongly with narcissism than did IM, $z = 5.67$, *p* < .001. In line with previous research (Paulhus, 1998), we found evidence that SDE and IM show differential relations with self-enhancement. Furthermore, the BIDR-16's pattern of correlations mirror those of the BIDR-40 (Table 3): SDE from each version correlated with self-esteem equally strongly, $z = 1.34$, *ns*, and IM from each version correlated with self-esteem equally strongly, $z = -1.78$, *ns*. Moreover, SDE from the BIDR-40 and BIDR-16 correlated with narcissism equally, $z = 1.75$, *ns*, and IM from each version correlated with narcissism equally, $z = 0.34$, *ns*. Thus, our item refinement preserved the meaning and utility of SDE and IM in relation to self-enhancement.

Study 2

Study 2 attempted to replicate the CFA findings obtained in Study 1 using an independent sample. Study 1 relied on administration of the BIDR-40; however, in Study 2, we administered the BIDR-16 alone in an attempt to validate further the brief version.

Table 3. Correlations Between the BIDR-40, BIDR-16, and Existing Scales in Study 1.

Scale	IM-40	SDE-40	IM-16	SDE-16
SDE-40	.26***	—	—	—
IM-16	.84***	.32***	—	—
SDE-16	.28***	.87***	.32***	—
Self-esteem	.09***	.54***	.12***	.53***
Narcissism	-.02	.18***	-.03	.15***

Note. BIDR = Balanced Inventory of Desirable Responding; IM = impression management; SDE = self-deceptive enhancement. **p* < .05. ***p* < .005. ****p* < .001.

Method

Participants. Participants included 670 (487 women, one unidentified) online volunteers recruited via research websites (e.g., <http://psych.hanover.edu/research/exponnet.html>); $M_{\text{age}} = 29.43$, $SD = 12.62$, range = 16–70. They were from the United Kingdom ($n = 343$), the United States ($n = 204$), Europe ($n = 44$), Australasia ($n = 33$), Canada ($n = 17$), Indian subcontinent ($n = 9$), East Asia ($n = 8$), Africa ($n = 5$), Central/South America ($n = 5$), and the Middle-East ($n = 2$), and included students ($n = 285$) and non-students ($n = 385$).

Materials and procedure. Participants completed self-report measures via the Internet without compensation. After providing demographic information, they completed the BIDR-16 (1 = *totally disagree*, 8 = *totally agree*).

Results and Discussion

We used CFA to assess the goodness of fit of our refined 16-item, two-factor model. The results suggest a close fit to the data: $\chi^2(103, N = 670) = 405.60$, GFI = .92, CFI = .90, RMSEA = .07 (90% CI = [.06, .08]), SRMR = .06. As in Study 1, a one-factor model evidenced an unacceptable fit: $\chi^2(104, N = 670) = 612.99$, GFI = .88, CFI = .83, RMSEA = .09 (90% CI = [.08, .10]), SRMR = .07; with a difference of 1 *df*, the $\chi^2_{\text{Difference}} = 207.39$, *p* < .001. Thus, the data suggest that the BIDR-16 more likely reflects two dimensions than a single SDR dimension.

Study 3

In Study 3, we examined the test–retest reliability of the BIDR-16 over a 2-week interval. Previous research using the BIDR-40 revealed test–retest correlations of $r = .69$ for SDE and $r = .65$ over 5 weeks (Paulhus, 1991).

Method

Participants. Participants included 352 (219 women) students from a University in the United Kingdom ($M_{\text{age}} = 20.46$, $SD = 3.80$, range = 20–51).

Materials and procedure. Participants completed several measures via the Internet at two time points, two weeks apart. After providing demographic information, participants completed measures in randomized order, including the BIDR-16 (1 = *totally disagree*, 8 = *totally agree*). Participants received £5 for participating.

Results and Discussion

Table 1 presents descriptive statistics for SDE and IM. Scores on the BIDR-16 were stable over a two week period, with test-retest reliability for SDE, $r = .79$, $p < .001$, and for IM, $r = .74$, $p < .001$, which are in the same order of magnitude as the BIDR-40 (Paulhus, 1991).

Study 4

The goal of Study 4 was to assess the construct validity of both BIDR-16 subscales using another independent sample. We examined the extent to which SDE and IM correlated with a commonly used SDR scale, self-enhancement indices, and Big Five personality traits.

Method

Participants. Participants included 708 (564 women) online volunteers, recruited via research websites as in Study 2 ($M_{\text{age}} = 22.30$, $SD = 8.26$, range = 16-74). Most were from the United States ($n = 477$) and the United Kingdom ($n = 174$), others from Canada ($n = 17$), Australasia ($n = 14$), Europe ($n = 14$), East Asia ($n = 6$), and other regions or undeclared ($n = 7$), and included students ($n = 588$) and non-students ($n = 120$).

Materials and procedure. Participants completed self-report measures via the Internet without compensation. After providing demographic information, they completed measures in randomized order. In addition to the BIDR-16 (1 = *totally disagree*, 8 = *totally agree*), participants completed scales for construct validation purposes.

We assessed SDR with a 10-item short form of the MCSDS (Strahan & Gerbasi, 1972, $\alpha = .55$).⁵ Items include, "I have never intensely disliked anyone" (*true/false*).

We assessed self-enhancement with the RSES ($\alpha = .89$; 1 = *strongly disagree*, 6 = *strongly agree*), the NPI-40 ($\alpha = .85$), and a brief version of the How-I-See-Myself scale (B-HSM; Campbell, Rudich, & Sedikides, 2002, $\alpha = .75$). Participants rated themselves (1 = *much less than the average person*, 6 = *much more than the average person*) on eight adjectives (e.g., *assertive, kind*).

We assessed personality traits with the Ten-Item Personality Inventory (Gosling, Rentfrow, & Swann, 2003), and report the correlation between each of the five-item pairs. Respondents rated themselves (1 = *strongly disagree*, 6 = *strongly agree*) on 10-trait pairs: Extraversion: for

Table 4. Correlations Between BIDR-16, SDE and IM, and Existing Scales in Study 4.

Scale	IM	SDE	z
SDE	.34***	—	—
MCSDS	.53***	.32***	5.48***
Self-esteem	.18***	.52***	-8.61***
Narcissism	-.18***	.26***	-10.14***
B-HSM	.10*	.26***	-3.82**
Extraversion	-.06	.16***	-4.96***
Emotional Stability	.26***	.45***	-4.81***
Openness	.04	.18***	-3.35**
Conscientiousness	.20***	.25***	-1.34
Agreeableness	.37***	.13**	5.89***

Note. BIDR = Balanced Inventory of Desirable Responding; SDE = self-deceptive enhancement; IM = impression management; MCSDS = Marlowe-Crowne Social Desirability Short; B-HSM = Brief How-I-See-Myself.

* $p < .05$. ** $p < .005$. *** $p < .001$.

example, *extraverted, enthusiastic*, $r = .47$, $p < .001$; Emotional stability: for example, *anxious, easily upset*, $r = .52$, $p < .001$; Openness: for example, *open to new experiences, complex*, $r = .22$, $p < .001$; Conscientiousness: for example, *dependable, self-disciplined*, $r = .39$, $p < .001$; Agreeableness: for example, *critical, quarrelsome*, $r = .11$, $p < .005$.

Results and Discussion

SDR. The BIDR-16 index of SDE correlated modestly with IM (Table 4), with equal magnitude to Study 1, $z = -0.51$, *ns*. Table 4 shows that the MCSDS correlated more strongly with IM than with SDE, consistent with Paulhus and Reid (1991).

Self-enhancement. The present study evidenced similar positive correlations to those in Study 1 between SDE and self-esteem, $z = 0.31$, *ns*, and IM and self-esteem, $z = -1.40$, *ns*, with SDE correlating more strongly with self-esteem than IM. SDE correlated positively with narcissism (Table 4), as in Study 1 but more strongly, $z = -2.86$, $p = .004$. IM correlated negatively with narcissism, as in Study 1 but more strongly, $z = 3.46$, $p < .001$, mirroring the pattern of results found by Borkenau and Zaltauskas (2009). Finally, the correlation between B-HSM and SDE was larger than the correlation between B-HSM and IM (Table 4). Thus, all self-enhancement indices related positively to SDE but weakly or negatively to IM.

Personality traits. SDE correlated positively and most strongly with emotional stability, followed by conscientiousness, openness, extraversion, and agreeableness (Table 4). IM correlated positively and most strongly with agreeableness, followed by emotional stability and

conscientiousness, and did not correlate with extraversion or openness. Such relations are consistent with those reported by Li and Bagger (2006).

General Discussion

SDR continues to present a challenge to self-report measurement (Stöber et al., 2002). This ubiquitous problem has led to the development of many scales over the years to screen for biased responding. The most popular scale is the MCSDS (Crowne & Marlowe, 1964), despite criticisms of its low reliability, outdated wording, and unidimensional factor structure (e.g., Beretvas et al., 2002). Paulhus (1991, 1998) developed the BIDR-40, which captures the two-dimensional nature of SDR, and provides an important theoretical and empirical extension to SDR research; however, short forms of the MCSDS are often preferred because of the BIDR's length. The aim of this research was to create a shorter version of the BIDR, which is psychometrically equivalent, that is, retains the original scale's two-factor structure, reliability, and validity.

Accordingly, in Study 1, we evaluated the BIDR-40 on model fit and dimensionality and refined the original scale, reducing to 16-items while maintaining model fit. The resulting BIDR-16 displayed superior fit for a two-factor than a one-factor model, confirming it reflects two dimensions (SDE and IM). The two short form scales remained conceptually similar to those of the long form, as demonstrated by high correlations between the short and long form and by similar correlations of the long and short forms with external correlates. Study 2 replicated the CFA findings using an independent sample administering only the BIDR-16. Study 3 provided evidence for the temporal stability of the BIDR-16, and Study 4 provided further evidence of the validity of the BIDR-16, replicating previous relationships between the long form and measures of SDR, self-enhancement, and personality traits.

Controversy over the dimensionality of SDR is ongoing. Many researchers who use the BIDR-40 continue to calculate the two originally proposed subscales, IM and SDE, and in this respect, the BIDR-16 represents an excellent substitute for the long version. To the extent that SDR is best represented by a three-factor (Paulhus & Reid, 1991) or four-factor structure (Paulhus & Trapnell, 2009), then we can only claim to measure two of the three or four types of SDR.

In all, using large and relatively diverse samples, this research provides evidence that scores on the BIDR-16 are adequately reliable and valid; demonstrating this shortened scale is a reasonable substitute for the BIDR-40 in studies where length of assessment is a concern. With eight items per subscale, the BIDR-16 is short enough to reduce transient errors that may occur as a result of fatigue or boredom but long enough for participants to get into a suitable mind-set for responding to items. Although internal consistencies of the BIDR-16 are relatively low (i.e.,

not always exceeding .70), they are comparable with those of the BIDR-40 (Li & Bagger, 2007). Moreover, given that internal consistency indexes construct breadth (Clark & Watson, 1995), the BIDR's moderate internal consistency is a reflection that SDE and IM, respectively, entail a broad range of self-enhancement and IM instantiations. Importantly, the high-test-retest correlations of SDE and IM attest to their high reliability. The studies outlined here demonstrate the validity of our shortened scale. Future research using the BIDR-16 will continue to build its nomological network.

Many nations have recently started collecting large-scale and nationally representative data. These samples often ignore the important issue of SDR, perhaps in part because of scale length. We hope the BIDR-16 proves useful in this regard. We believe the BIDR-16 offers researchers advantages over previously available scales, making it more practical to assess validly and to control for both SDR dimensions.

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) received no financial support for the research and/or authorship of this article.

Notes

1. Dataset 1, $n = 97$ (online; measures Single-Item Self-Esteem measure [1 = *disagree strongly*, 11 = *agree strongly*] and Narcissistic Personality Inventory-40 [NPI-40]); Dataset 2, $n = 115$ (online; measures Single-Item Self-Esteem measure [1 = *disagree strongly*, 11 = *agree strongly*] and NPI-40); Dataset 3, $n = 198$ (online; measures Self-Liking Self-Competence Scale-Revised [1 = *strongly disagree*, 7 = *strongly agree*]); Dataset 4, $n = 187$ (paper and pencil; measures RSES [1 = *strongly disagree*, 11 = *strongly agree*]); Dataset 5, $n = 239$ (paper and pencil); Dataset 6, $n = 352$ (paper and pencil; measures RSES [1 = *strongly disagree*, 4 = *strongly agree*], NPI-40); Dataset 7, $n = 345$ (online; measures RSES [1 = *strongly disagree*, 4 = *strongly agree*], NPI-15); Dataset 8, $n = 415$ (online; measures RSES [1 = *strongly disagree*, 4 = *strongly agree*], NPI-15).
2. Participants in Datasets 1 to 6 completed BIDR Version 6, and participants in Datasets 7 to 8 completed BIDR Version 7.
3. Results available from the corresponding author.
4. The z scores reported are the result of Meng, Rosenthal, and Rubin's (1992) test of correlated correlation coefficients.
5. Consistent with previous criticisms of the MCSDS, this short form also suffered low reliability, with an $\alpha < .7$ (Cronbach, 1951). Furthermore, consistent with previous suggestions that the MCSDS is not a unidimensional measure of need for approval (e.g., Paulhus, 1984), we demonstrated that items from the short MCSDS correlated positively with both facets of SDR.

References

- Ballard, R. (1992). Short forms of the Marlowe–Crowne Social Desirability Scale. *Psychological Reports*, 71, 1155–1160. doi:10.2466/PRO.71.8.1155-1160
- Ballard, R., Crino, M. D., & Rubinfeld, S. (1988). Social desirability response bias and the Marlowe–Crowne Social Desirability Scale. *Psychological Reports*, 63, 227–237. doi:10.2466/pr0.1988.63.1.227
- Barger, S. (2002). The Marlowe–Crowne affair: Short forms, psychometric structure and social desirability. *Journal of Personality Assessment*, 79, 286–305. doi:10.1207/S15327752JPA7902_11
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606. doi:10.1037//0033-2909.88.3.588
- Beretvas, N. S., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe–Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 4, 570–589. doi:10.1177/0013164402062004003
- Borkenau, P., & Zaltauskas, K. (2009). Effects of self-enhancement on agreement on personality profiles. *European Journal of Personality*, 23, 107–123. doi:10.1002/per.707
- Brockway, J. H., Carlson, K. A., Jones, S. K., & Bryant, F. B. (2002). Development and validation of a scale for measuring cynical attitudes toward college. *Journal of Educational Psychology*, 94, 210–224. doi:10.1037/0022-0663.94.1.210
- Campbell, W. K., Goodie, A. S., & Foster, J. D. (2004). Narcissism, overconfidence, and risk attitude. *Journal of Behavioral Decision Making*, 17, 297–311. doi:10.1002/bdm.475
- Campbell, W. K., Rudich, E. A., & Sedikides, C. (2002). Narcissism, self-esteem, and the positivity of self-views: Two portraits of self-love. *Personality and Social Psychology Bulletin*, 28, 358–368. doi:10.1177/0146167202286007
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319. doi:10.1037/1040-3590.7.3.309
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. doi:10.1007/bf02310555
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York, NY: Wiley.
- Damarin, E., & Messick, S. (1965). *Response styles as personality variables: A theoretical integration of multivariate research* (Research Bulletin No. 65-10). Princeton, NJ: Educational Testing Service.
- Gignac, G. E. (2013). Modeling the Balanced Inventory of Desirable Responding: Evidence in favor of a revised model of socially desirable responding. *Journal of Personality Assessment*, 95, 645–656. doi:10.1080/00223891.2013.816717
- Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, 11, 340–344. doi:10.1111/j.0965-075X.2003.00256.x
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528. doi:10.1016/S0092-6566(03)00046-1
- Greenberger, E., Chen, C., Dmitrieva, J., & Farrugia, S. P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, 35, 1241–1254. doi:10.1016/S0191-8869(02)00331-8
- Herbert, J. R., Clemow, L., Pbert, L., Ockene, I. S., & Ockene, J. K. (1995). Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *International Journal of Epidemiology*, 24, 389–398. doi:10.1093/ije/24.2.389
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453. doi:10.1037//1082-989X.3.4.424
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford Press.
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). London, England: SAGE.
- Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.8 user's reference guide*. Uppsala, Sweden: Scientific Software International.
- Kam, C. (2013). Probing item social desirability by correlating personality items with Balanced Inventory of Desirable Responding (BIDR): A validity examination. *Personality and Individual Differences*, 54, 513–518. doi:10.1016/j.paid.2012.10.017
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in Structural Equation Modeling. *Structural Equation Modeling*, 10, 333–351. doi:10.1207/S15328007SEM1003_1
- Klesges, L. M., Baranowski, T., Beech, B., Cullen, K., Murray, D. M., Rochon, J., & Pratt, C. (2004). Social desirability bias in self-reported dietary, physical activity and weight concern measures in 8- to 10-year-old African-American girls: Results from the Girls health Enrichment Multisite Studies (GEMS). *Preventive Medicine*, 38, 78–87. doi:10.1016/j.ypmed.2003.07.003
- Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity of scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement*, 67, 859–876. doi:10.1177/0013164406299104
- Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe–Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement*, 65, 140–154. doi:10.1177/0013164404267285
- Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 131–141. doi:10.1111/j.1468-2389.2006.00339.x
- Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A reliability generalization study. *Educational and Psychological Measurement*, 67, 525–544. doi:10.1177/0013164406292087
- McIntire, S. A., & Miller, L. A. (2000). *Foundations of psychological testing*. New York, NY: McGraw-Hill.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, 111, 172–175. doi:10.1037/0033-2909.111.1.172
- Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effect of social desirability response sets in organizational behavior research. *Journal of Occupational and Organizational Psychology*, 56, 131–149. doi:10.1111/j.2044-8325.1992.tb00490.x

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609. doi:10.1037/0022-3514.46.3.598
- 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.
- Paulhus, D. L. (1994). *Balanced inventory of desirable responding: Reference manual for BIDR version 6*. Unpublished manuscript, University of British Columbia, Vancouver, Canada.
- Paulhus, D. L. (1998). *Manual for the Paulhus Deception Scales: BIDR Version 7*. Toronto, Ontario, Canada: Multi-Health Systems.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67-88). Hillsdale, NJ: Lawrence Erlbaum.
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 60, 307-317. doi:10.1037//0022-3514.60.2.307
- Paulhus, D. L., & Trapnell, P. D. (2009). Self-presentation of personality: An agency- communion framework. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality psychology* (pp. 493-517). New York, NY: Guilford Press.
- Raskin, R., & Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validation. *Journal of Personality and Social Psychology*, 54, 890-902. doi:10.1037/0022-3514.75.1.219
- Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social Desirability Scale. *Journal of Clinical Psychology*, 38, 110-125. doi:10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 27, 151-161. doi:10.1177/0146167201272002
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Sackeim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139-197). New York, NY: Plenum Press.
- Schmidt, F. L., Le, H., & Iles, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8, 206-224. doi:10.1037/1082-989X.8.2.206
- Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt: Psychometrische Eigenschaften einer Lang- und einer Kurzform des Deutschen NPI [Measuring narcissism as a personality construct: Psychometric properties of a long and a short version of the German Narcissistic Personality Inventory]. *Diagnostica*, 50, 202-218.
- Steenkamp, J.-B. E. M., de Jong, M. G., & Baumgartner, H. (2010). Socially desirable response tendencies in survey research. *Journal of Marketing Research*, 47, 199-214. doi:10.1509/jmkr.47.2.199
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17, 222-232. doi:10.1027//1015-5759.17.3.222
- Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment*, 78, 370-389. doi:10.1207/S15327752JPA7802_10
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogenous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 28, 191-193. doi:10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G
- Tafarodi, R. W., & Swann, W. B. (2001). Two-dimensional self-esteem: Theory and measurement. *Personality and Individual Differences*, 31, 653-673. doi:10.1016/S0191-8869(00)00169-0
- Thompson, E., & Phua, F. (2005). Reliability among senior managers of the Marlowe-Crowne Short Form Social Desirability Scale. *Journal of Business and Psychology*, 19, 541-554. doi:10.1007/s10869-005-4524-4
- Uziel, L. (2010). Rethinking social desirability scales: From impression management to interpersonally oriented self-control. *Perspectives on Psychological Science*, 5, 243-262. doi:10.1177/1745691610369465

Author Biographies

Claire M. Hart is a lecturer in Personality/Social Psychology at the University of Southampton, UK. Her primary research interests lie in the interpersonal deficits of narcissists and the ways in which narcissists maintain their positive self-views.

Timothy D. Ritchie is an associate professor of psychology and department chairperson at Saint Xavier University in Chicago, IL, USA. His primary research interests include the relations between autobiographical memory, the self, and emotions, and the relations between stress, the self, and subjective well-being.

Erica G. Hepper is a lecturer in Personality/Social Psychology at the University of Surrey, UK. Her research examines the interplay between the individual and relational self, including work on adult attachment, nostalgia, narcissism, and self-evaluation motives.

Jochen E. Gebauer is the head of the Emmy Noether junior research group "Self & Society" at the University of Mannheim. His research concerns the self-concept from the perspectives of personality and social psychology.