Validity of the SDS-17 measure of social desirability in the American context

Brian F. Blake a, Jillian Valdiserri b, Kimberly A. Neuendorf c,*, Jacqueline Nemeth d

a Department of Psychology, Cleveland State University, Cleveland, OH 44115, United States
b Marketing Strategies Inc., Livonia, MI 48152, United States
c School of Communication, Cleveland State University, 2121 Euclid Avenue, MU 241, Cleveland, OH 44115, United States
d American Greetings, Cleveland, OH 44144, United States

Received 26 July 2005; received in revised form 1 December 2005; accepted 7 December 2005
Available online 9 March 2006

Abstract

In Germany, Stober et al. (1999, 2001) presented evidence for the validity of the SDS-17, a new measure of social desirability bias. In the current investigation, three experiments \((n = 800)\) assessed the SDS-17’s validity in the US environment. In all conditions SDS-17 scores correlated highly with Marlowe–Crowne scores. In Study 1, a group administration of a paper and pencil booklet, SDS-17 scores of 327 college students were higher under Fake Good than Standard conditions, and both were higher than scores in the Honest condition. Study 2, an online survey of a demographically diverse adult sample \((n = 257)\), showed that the increase in SDS-17 scores under Fake Good conditions occurs also in a Web survey and that SDS-17 scores were unrelated to one’s demographic profile. Study 3, a group administration to 216 college students, revealed again that scores under Fake Good were higher than those under Standard administration and that SDS-17 scores correlated more highly with the Impression Management than with the Self-Deception subscales of the BIDR. The SDS-17 appeared valid for the US environment as a measure of socially desirable responding. The evidence, however, encourages its further assessment as an index of social desirability bias per se.

© 2006 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +1 216 687 3994; fax: +1 216 687 5435.
E-mail address: k.neuendorf@csuohio.edu (K.A. Neuendorf).
1. Introduction

Behavioral scientists (e.g., Crowne & Marlowe, 1960; Edwards, 1957) have long been concerned with social desirability bias (SDB), defined as distorting one’s self-presentation to make a favorable impression upon others. It continues to be seen in various quarters, for example, consumer psychology (e.g., Fisher, 2000), as a source of contamination that should be controlled in self-report measures. It has been well established (e.g., Barger, 2002; Ellingson, Sackett, & Hough, 1999; Helmes & Holden, 2003), though, that SDB may be but a component of the more general multidimensional phenomenon of socially desirable responding (SDR), defined as presenting oneself as having characteristics appreciated by others. Beyond reflecting just SDB, SDR may represent the internalization of cultural values (Fisher & Katz, 2000), the expression of ongoing personality traits like emotional stability and conscientiousness (Ones, Viswesvaran, & Reiss, 1996), or an overly favorable but honest self-evaluation devised to maintain self-esteem (Paulhus, 1984), among other factors.

Attempts to unravel the role of SDB face the problem that there is no consensually agreed-upon measure of SDB. One of the closest constructs is the Impression Management dimension of Paulhus’s Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984, 1994, 2002); this refers to the enhanced positivity of one’s self-presentation resulting from situational demands or temporary inclinations to describe oneself to others in favorable terms. The conceptual basis for this measure, though, has been questioned (e.g., Helmes & Holden, 2003; Kroner & Weekes, 1996; Pauls & Crost, 2004). A new measure of SDB with demonstrated construct validity could be helpful in clarifying the nature and impacts of SDB.

In Germany, Stober (1999, 2001) proposed a new measure of social desirability bias. It was intended to overcome particular limitations of the most used index of SDR (e.g., Beretvas, Meyers, & Leite, 2002), the Marlowe–Crowne Scale (Crowne & Marlowe, 1960) and its shorter subscales (e.g., Barger, 2002). Initial validation studies (Stober, 1999, 2001) and applications (Stober & Wolfradt, 2001) were strongly supportive. However, as is true for psychological measures in general (e.g., Van de Vijver, 2003), the validity of SDR measures may not be directly generalizable across cultures (e.g., Johnson & Van de Vijver, 2003). The purpose of this series of studies was to assess the reliability and validity of the SDS-17 in the United States context.

1.1. SDS-17 development

As conceived by Stober (1999, 2001), “social desirability” is a readiness to give biased, distorted self-descriptions that portray oneself in a manner that can make a favorable impression on others. The SDS-17 is composed of 16 true–false items, e.g., “I will never live off other people,” “I sometimes litter.” It is a balanced index in that one’s score is increased by a true response on nine items, and by a false response on seven items. Its statements were intended to have contemporary referents and phrasing. Stober observed that various items on the original and shortened forms of the Marlowe–Crowne (MC) are dated (e.g., “My table manners at home are as good as when I eat
out in a restaurant;” “I never make a long trip without checking the safety of my car”). Indeed, similar observations have been made for quite some time (e.g., Ballard, Crino, & Rubenfeld, 1988). Further, MC statements often incorporate culture bound referents (e.g., driving cars, voting for political candidates), as do other SDR scales such as the BIDR of Paulhus (1984, 1994, 2002). The SDS-17 contains few obvious references to cultural institutions and organizations. Moreover, since the length of the MC (33 items) can be troublesome (e.g., Fisher, 2000), the brevity of the SDS-17 is advantageous.

Convergent validity was suggested (Stober, 1999, 2001) by the scale’s correlation with the MC (.74, \( p < .01 \), in the 1999 and .68, \( p < .01 \), in the 2001 study), the revised Lie Scale of the Eysenck Personality Questionnaire (\( r = .60, p < .01 \); Eysenck & Eysenck, 1991), and the Sets of Four Scale (\( r = .52, p < .01 \); Borkenau & Ostendorf, 1992). Discriminant validity was indicated by non-significant correlations with neuroticism, extraversion, psychoticism, and with openness to experience as gauged by the revised Eysenck Personality Questionnaire and the NEO Five Factor Inventory (Costa & McCrae, 1993), respectively. Discriminant validity was lower than would be desirable for an index intended to measure SDB, though, as seen in significant positive correlations with agreeableness and conscientiousness measured by the Five Factor Inventory (Costa & McCrae, 1993).

Befitting an index intended to assess readiness to provide distorted responses, Stober (2001) demonstrated that SDS-17 scores were lower when respondents were given “Standard” questionnaire instructions (i.e., told to complete the questionnaire truthfully) than when given instructions to produce responses presenting oneself favorably (i.e., told to imagine that the way they responded was important to their application for a good job, typically called a “Fake Good” condition). Further, comparison was made with the two dimensions of the BIDR (version 6, Paulhus, 1994; German version, Musch, 1999). The Impression Management (IM) subscale of the BIDR was developed to reflect conscious dissimulation of responses with the aim of making a favorable impression on others, while the Self-Deceptive Enhancement (SDE) subscale referred to responding geared to seeing oneself in a favorable light and so to protect positive self-esteem (Paulhus, 1984). Stober (2001) found that the SDS-17 correlated higher with the IM subscale (\( r = .37, p < .01 \)) than with the SDE subscale (\( r = .21, p < .05 \)), suggesting that the SDS-17 represented distorted self-presentation more than a chronic tendency to protect one’s self-esteem via thinking of oneself as having appealing characteristics.

1.2. Unanswered questions

1.2.1. National setting

Is the SDS-17 scale appropriate for use in the United States? Responses that convey a favorable impression in one society do not necessarily do so in another (e.g., Middleton & Jones, 2000; Smith, Smith, & Seymour, 1993), and validation may not transfer directly across nations (cf. Van de Vijver, 2003). Hence, the SDS-17 must be empirically validated for use in the US and other nations.

1.2.2. Motivation to distort

In the initial validation research (Stober, 1999, 2001), subject responses in the assumed “Fake Good” (job application) condition were compared to those in the “Standard” (respond frankly)
condition. Two issues can be raised about this procedure. First, it is possible that one may “Fake
Good” differently in a job interview than in other settings (e.g., applying for fraternity membership or meeting a blind date). A more conservative approach, then, would be to instruct subjects
to make themselves appear good or appealing in the eyes of others without specifying a particu-
lar context. Second, under the Standard conditions it is quite possible that subjects are still
motivated to distort. A more convincing approach would be to establish, in addition to the Stan-
dard setting, an “Honest” condition in which subjects are strongly encouraged to respond
truthfully.

1.2.3. Administration vehicle
Given the increasing popularity of internet surveys in the US, is the SDS-17 valid not only in
the traditional group administration via paper and pencil booklets but also in the online environ-
ment? Differences in SDR can occur between online and paper and pencil (e.g., Pettit, 2002), and
between responding alone or in a group (e.g., Richman, Kiesler, Weisbad, & Drasgow, 1999).
Although such studies have not yielded entirely consistent results (Hancock & Flowers, 2001),
a conservative approach would be to validate the SDS-17 separately for group paper and pencil
and for online administration.

1.2.4. Demographic differences
In the US, do SDS-17 scores vary with one’s demographic characteristics? Stober (2001) found
in Germany that scores increased with age, but were unrelated to gender. In the US are there dif-
fferences associated not only with age and gender but also with other key variables, notably edu-
cation, employment status, income, or marital status?

1.3. Overview of studies
Study 1, based on group administration of a print questionnaire to a university student sample,
asked whether SDS-17 scores are higher when individuals attempt to answer in a socially desirable
manner (Fake Good condition) than when responding in a typical test administration (Standard
condition) and whether both are higher than when respondents are asked to respond with frank
and open answers (Honest condition). Study 2 asked whether the increase in SDS-17 scores under
Fake Good conditions occurs among post-college-age adults taking part in an online survey. Both
Studies 1 and 2 assessed the SDS-17’s convergent validity by asking whether it correlates with the
MC scale. In Study 3, university students again responded to a print questionnaire in a group
administration. Included in the booklet were the IM and SDE subscales of the Paulhus BIDR
(1994) as well as the MC scale. Again, responses were obtained under two conditions, Fake Good
and Standard.

1.4. Predictions
Validity of the SDS-17 as an index of SDR in general and SDB in particular would be indicated
if, first, in all three studies the SDS-17 scores were higher under Fake Good than under Standard
conditions and if, second, in Study 1 scores in both the Fake Good and Standard conditions were
higher than those in the Honest condition. Third, validity as an index of SDR (but not necessarily
of SDB) would be suggested if SDS-17 scores correlated positively with Marlowe–Crowne scores in all conditions in all three studies, although the correlations under Fake Good conditions may be less reflective of the scale’s construct validity than would be the correlations in the Standard or Honest conditions (cf. Pauls & Crost, 2004). Fourth, a stronger correlation of the SDS-17 with the Impression Management than with the Self-Deceptive Enhancement scales would support the Stober conceptualization of the SDS-17 as an index of SDB. Fifth, while there was no compelling basis for offering a prediction, it was asked whether SDS-17 scores varied systematically with one’s demographic characteristics.

1.5. Selection of validation measures

The MC scale, 33 true–false items balanced for positive and negative wording, was used to establish convergent validity of the SDS-17 as a measure of the SDR because the MC is the most widely employed indicator of SDR (Beretvas et al., 2002). Due to its multidimensionality (e.g., Ellingson et al., 1999; Helmes & Holden, 2003), among other factors (Barger, 2002), the MC cannot be construed as a measure purely of SDB.

Despite challenges to the BIDR’s validity as an index of SDB (e.g., Helmes & Holden, 2003), the version 6 IM and SDE subscales were used because (a) the IM dimension represents the construct underlying the SDS-17, as demonstrated by Stober’s (2001) use of it to assess the construct validity of the SDS-17; (b) the IM–SDE distinction is used in numerous settings (cf. Ellingson et al., 1999); (c) a body of some research does support its assumptions (e.g., Hancock & Flowers, 2001; Musch, Brockhaus, & Broeder, 2002); and (d) as noted earlier, there is a dearth of measures of the SDB construct that have been widely used and successfully validated as an index uniquely measuring SDB. Both the 10 item IM subscale (e.g., “I sometimes tell lies if I have to”) and the 10 item SDE subscale (e.g., “I have always been honest with myself”) have a true–false response format and are balanced for positively and negatively worded statements.

2. Method

2.1. Study 1

2.1.1. Subjects

Undergraduate students (n = 327) enrolled in psychology and communication courses at a larger midwestern urban university (n = 252) and at a smaller residential college (n = 75) were offered course credit for participation in a survey of their shopping behavior and of their orientation toward a variety of issues.

2.1.2. Conditions

Subjects were randomly assigned to one of three conditions and were given a questionnaire booklet. In the Standard condition (conducted in the manner suggested by Dillman (2000) for group administration of a print questionnaire), subjects were told that their answers would be confidential and they were asked to respond truthfully. In the Honest condition, it was explained that the experiment was designed to evaluate how the questionnaire was susceptible to distortion
by persons attempting to make themselves appear attractive in the eyes of others. It was noted that they had been randomly assigned to the Honest condition and that other persons were being asked to “Fake Good.” Investigators exhorted the subjects to answer as frankly as possible and assured them their answers would be confidential. In the Fake Good condition, subjects were also informed that the goal of the study was to test the responsivity of the questionnaire to social desirability bias. They were told that they were not to answer in terms of their actual opinions but to answer in a manner which would make them appear to be a good person with admirable qualities, without specifying which qualities they should emphasize. Subjects were instructed to answer the demographic questions truthfully.

2.1.3. Questionnaire

The booklet included the 16 SDS-17 and the 33 MC items, along with the six demographic characteristics described in Table 1. Other items, irrelevant to this study, pertained to shopping for products and services.

Table 1
Subject characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>327</td>
<td>257</td>
<td>216</td>
</tr>
<tr>
<td>Fake Good</td>
<td>115</td>
<td>121</td>
<td>105</td>
</tr>
<tr>
<td>Standard</td>
<td>118</td>
<td>136</td>
<td>111</td>
</tr>
<tr>
<td>Honest</td>
<td>94</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Age (mean years)</td>
<td>21.57</td>
<td>36.45</td>
<td>22.95</td>
</tr>
<tr>
<td>Gender (% male)</td>
<td>49.7</td>
<td>36.2</td>
<td>33.0</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>–</td>
<td>6.7%</td>
<td>–</td>
</tr>
<tr>
<td>Post-high school</td>
<td>93.4%</td>
<td>23.3%</td>
<td>90.3%</td>
</tr>
<tr>
<td>College graduate</td>
<td>6.4%</td>
<td>70.0%</td>
<td>9.7%</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Full-time</td>
<td>17.1</td>
<td>35.8</td>
<td>12.5</td>
</tr>
<tr>
<td>% Part-time</td>
<td>45.6</td>
<td>13.2</td>
<td>48.6</td>
</tr>
<tr>
<td>Other(^a)</td>
<td>37.3%</td>
<td>51.0%</td>
<td>38.9%</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–$20,000</td>
<td>27.2%</td>
<td>12.7%</td>
<td>36.9%</td>
</tr>
<tr>
<td>$20,001–$30,000</td>
<td>22.0%</td>
<td>8.6%</td>
<td>10.7%</td>
</tr>
<tr>
<td>$30,001–$50,000</td>
<td>5.8%</td>
<td>23.0%</td>
<td>17.0%</td>
</tr>
<tr>
<td>$50,001 or more</td>
<td>38.5%</td>
<td>55.8%</td>
<td>35.4%</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>89.9%</td>
<td>39.6%</td>
<td>86.6%</td>
</tr>
<tr>
<td>Married</td>
<td>5.5%</td>
<td>52.9%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Divorced, widowed, separated</td>
<td>4.0%</td>
<td>7.5%</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

\(^a\) Includes unemployed, homemakers, retired, etc.
2.2. Study 2

2.2.1. Respondents

A convenience sample of 257 friends, relatives, coworkers, and others known to graduate students on the research team were asked to participate. An attempt was made to secure as demographically and geographically diverse a sample as possible as long as respondents were over 21 years of age. When recruited, participants were told simply that the survey involved measuring their orientations and consumer behavior and that their answers would be confidential. Participants were directed to the University’s website containing the questionnaire.

2.2.2. Conditions

When accessing the website, respondents were assigned randomly to the Standard or the Fake Good conditions and given the same instructions and explanations as were used in Study 1.

2.2.3. Questionnaire

The instrument contained the same measures as did Study 1, but formatted for online presentation.

2.3. Study 3

The 216 subjects were recruited from psychology and communication undergraduate classes at a large midwestern urban university and assigned to either the Standard or the Fake Good condition in the same manner as in Study 1. This questionnaire also contained the 20 BIDR items.

2.4. Subject characteristics

Of 814 participants, 800 provided usable data (see Table 1). Study 1 was composed of young, mainly single students, a substantial number of whom were employed part-time. There was considerable spread in their own or family income. Study 2 respondents were predominantly well educated and fairly affluent. The student sample in Study 3 was again generally young, single, and employed at least part-time. Both Studies 2 and 3 respondents were predominantly female. Thus, the two student samples were fairly typical of many US student populations used in psychological research and the online sample, while upscale, included more demographic diversity.

3. Results

3.1. Preliminary analyses

3.1.1. Demographic differences between conditions

Differences between conditions in respondent profiles on the seven demographic characteristics were assessed separately for each study. Only three of the 35 comparisons were statistically significant and only one (income in Study 2) was of any magnitude. In light of the lack of relationship
in Study 2 between respondent demographic characteristics and SDS-17 scores as discussed later, these differences were felt to be inconsequential.

3.1.2. Homogeneity of the validation scales

Table 2 presents the KR-20 coefficients for the MC and the BIDR scales in each condition. The two scales appeared to be highly reliable in each condition in each study and so could serve as evidence of convergent validity for the SDS-17 measure.

3.1.3. Establishment of conditions

Tests of differences among the conditions in scores on the MC Scale revealed the following: In Study 1, an ANOVA indicated that the conditions differed significantly ($F_{2,313} = 11.67, p < .001$). The Tukey HSD showed that MC scores in the Fake Good condition ($M = 24.49, n = 113$) were higher than in the Standard condition ($M = 16.34, n = 116, p < .001$) and the Honest condition ($M = 13.99, n = 87, p < .001$), while scores for the Standard were higher than those for the Honest group ($p < .01$). In Study 2, the MC scores were higher under Fake Good ($M = 25.13, n = 114$) than under Standard conditions ($M = 16.92, n = 123, t_{(235)} = -12.72, p < .001$). In Study 3, the Levene test revealed that the variances were not homogenous within each group ($F = 7.75, p < .01$). The $t$-test (assuming heterogeneous variances) again found differences in mean MC scores between the Fake Good ($M = 30.84, n = 100$) and the Standard groups ($M = 15.64, n = 106, t_{(201.18)} = -23.05, p < .001$). It can be concluded that manipulation of the Fake Good, Standard, and Honest conditions was successful in all three studies, as verified via the MC, independently of the SDS-17 scores.

3.2. SDS-17 scale homogeneity

As shown in Table 2, the SDS-17 scale was highly consistent internally when individuals were attempting to Fake Good. Under more typical survey conditions the measure had satisfactory reliability in three of the four conditions. Overall, the measure appears to have satisfactory to good internal consistency.

3.3. Demographic differences

Under “normal” conditions do scores on the SDS-17 scale vary with one’s demographic profile? Using data from the Standard condition of Study 2, the demographic factors were converted to dummy variables and entered simultaneously as predictors in an ordinary least squares multiple

<table>
<thead>
<tr>
<th>Condition</th>
<th>Study 1</th>
<th>Study 2</th>
<th>Study 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MC</td>
<td>SDS-17</td>
<td>MC</td>
</tr>
<tr>
<td>Fake Good</td>
<td>.94</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>Standard</td>
<td>.76</td>
<td>.70</td>
<td>.71</td>
</tr>
<tr>
<td>Honest</td>
<td>.77</td>
<td>.76</td>
<td>–</td>
</tr>
</tbody>
</table>
regression against SDS-17 scores as the criterion. No significant differences between demographic groups emerged in their SDS-17 scores ($R = .32$, $R^2 = .10$, adjusted $R^2 = -.03$; $F_{16,106} = .76$, $p = .73$).

3.4. Response to conditions

The means and standard deviations of SDS-17 scores in each condition in each study are shown in Table 3.

In Study 1, an ANOVA showed that the differences among the conditions were significant ($F_{2,319} = 102.56$, $p < .001$). The Tukey HSD test observed a difference in the Fake Good vs. the Standard ($p < .001$) and vs. the Honest ($p < .001$), and in the Standard vs. the Honest ($p < .05$). Cohen’s $d$ showed the Standard–Fake Good difference ($d = 1.49$) and the Honest–Fake Good difference ($d = 1.79$) to be quite strong. The Standard–Honest difference ($d = .36$) was more modest, but meaningful. In Study 2, the difference was again significant in the predicted direction ($t_{(245)} = -13.66$, $p < .001$) and, given $d = 1.74$, quite powerful. Finally, in Study 3, the Levene test ($F = 28.74$, $p < .001$) showed that the within group variances were different. The $t$-test with heterogeneous variances was yet again significant ($t_{(195,00)} = -21.69$, $p < .001$) and the difference substantial ($d = 2.99$). Thus, in all three studies, individuals’ SDS-17 scores varied with conditions in the manner expected.

3.4.1. MC and BIDR scale correlations

Assessments of convergent validity were quite positive. In each condition in each study the Pearson product moment correlation between the SDS-17 scale and the MC was strong and positive (see Table 4). Table 4 also shows that the correlation between the SDS-17 index and
the overall BIDR was significant in each condition in Study 3. Additionally, the SDS-17 measure correlated more highly with the IM than with the SD subscale both in the Fake Good (Hotelling–Williams $t = -2.57, p < .05$) and in the Standard (Hotelling–Williams $t = -2.39, p < .05$) conditions.

4. Discussion

The three studies strongly support the viability of the SDS-17 as a measure of SDR: First, internal consistency was acceptable in six of the seven assessments; the exception (alpha = .63) might have reflected context effects in the questionnaire’s item sequence. Second, in all three experiments SDS-17 scores consistently increased with situational demands to display socially desirable responses. Third, convergent validity was established via substantial correlations with the Marlowe–Crowne and the BIDR scales, well-established SDR indicators. Fourth, internal consistency, known group and convergent validity were shown both for the online survey situation as well as for paper and pencil group administration. Fifth, the absence of a significant relationship with the six demographic characteristics in Study 2 renders the SDS-17 easier to use in practice, since a person’s demographic profile would not have to be considered when interpreting SDS-17 scores.

Whether SDS-17 scores represent solely SDB, however, is a separate question. That the scale does reflect SDB is supported by its higher correlation in Study 3 with the IM than with the SDE subscale of the BIDR. Further, in the Standard condition, SDS-17 correlated significantly with the IM, but not with the SDE, subscale. The SDS-17 did correlate low, but significantly, with the SDE scale in the Fake Good context. The latter correlation perhaps may be artifactual, given that conceptually unrelated scales (Ellingson et al., 1999) in general and social desirability measures (Pauls & Crost, 2004) in particular may intercorrelate more strongly under Fake Good conditions. The BIDR/SDS-17 correlations tend to support SDS-17 as an index of SDB, at least under typical administrative conditions. The caveat, however, arises from the dearth of a commonly agreed-upon standardized measure of SDB per se to serve as a basis for convergent validity.

Given the encouraging results obtained here, additional research is warranted to address issues beyond the limits of the present investigation. First, evidence for the construct validity of the SDS-17 or of any measure proposed as an index of SD bias (dissimulation, distortion) could profit from a rather infrequently used approach, the direct measurement of the magnitude of each person’s dissimulation/distortion on an item-by-item, scale-by-scale basis. For example, Ellingson et al. (1999) directly gauged the degree of dissimulation on an item via a repeated measures design comparing a person’s response under Fake Good and under Honest instructional conditions. If such analyses can be corrected for acquiescence (e.g., Van de Vijver, 2003), non-comparability of positively and negatively worded items (Wong, Rindfleisch, & Burroughs, 2003), sequence effects and other potential artifacts, they might help differentiate between SDB and SDR (particularly, non-SDB components of SDR). Second, norms based on a more nationally representative respondent sample would help investigators assess the degree of SDB/SDR shown by an individual or by a respondent sample as a whole (as embodied, for example, in the group mean).
Cross-cultural research requires measures that are valid within each relevant culture and, even better, that are equivalent across cultures (cf. Van de Vijver, 2003). Some support for the SDS-17’s validity is now available in the US as well as in Germany. If subsequent validation research is also supportive, the SDS-17 may be valuable for cross-cultural investigations.

References


Musch, J. (1999). German version of the balanced inventory of desirable responding (BIDR, version 6). Unpublished manuscript, University of Bonn, Germany.


