

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

The Social Desirability Scale-17 (SDS-17):
Convergent Validity, Discriminant Validity, and Relationship with Age

Joachim Stöber

Martin Luther University of Halle-Wittenberg

Address correspondence to:

Dr. Joachim Stöber
Martin Luther University of Halle-Wittenberg
Department of Educational Psychology
Franckesche Stiftungen, Haus 5
D-06099 Halle (Saale)
Germany
Phone: +49 345 5523789
Fax: +49 345 5527244
E-mail: stoeber@paedagogik.uni-halle.de

Summary. Four studies investigating the convergent validity, discriminant validity, and relationship with age of the Social Desirability Scale-17 (SDS-17) are presented. Where convergent validity is concerned, SDS-17 scores showed correlations between .52 and .85 with other measures of social desirability (Eysenck Personality Questionnaire-Lie Scale, Sets of Four Scale, Marlowe-Crowne Scale). Moreover, scores were highly sensitive to social-desirability provoking instructions (job-application instruction). Finally, with respect to the Balanced Inventory of Desirable Responding, SDS-17 scores showed a unique correlation with impression management, but not with self-deception. Where discriminant validity is concerned, SDS-17 scores showed nonsignificant correlations with neuroticism, extraversion, psychoticism, and openness to experience, whereas there was some overlap with agreeableness and conscientiousness. With respect to relationship with age, the SDS-17 was administered in a sample stratified for age, with age ranging from 18 to 89 years. In all but the oldest age group, the SDS-17 showed substantial correlations with the Marlowe-Crowne Scale. The influence of age (cohort) on mean scores, however, was significantly smaller for the SDS-17 than for the Marlowe-Crowne Scale. In sum, results indicate that the SDS-17 is a reliable and valid measure of social desirability, suitable for adults of 18 to 80 years of age.

Keywords: Social desirability, impression management, conscientiousness, instruction manipulation, age

Introduction

Forty years ago, Crowne and Marlowe (1960) developed a new scale for the measurement of social desirability because they were dissatisfied with the then predominant scale of Edwards (1953, 1957). The reason for this dissatisfaction was that Edwards had constructed his scale with items from the Minnesota Multiphasic Personality Inventory (MMPI). Because the MMPI mainly captures psychopathological aspects of personality, scores from Edwards' scale confounded social desirability with absence of psychopathology. For their new scale, Crowne and Marlowe (1960) therefore chose a different approach. First they selected items from personality questionnaires that described behaviors which they considered socially desirable but infrequent or socially undesirable but frequent, disregarding items with psychopathological implications. Then a sample of college students rated the degree of social desirability of each of the selected items. By retaining only items with high ratings and satisfactory item-total correlations, the authors arrived at a "new social desirability scale independent of psychopathology," now well known as the Marlowe-Crowne Scale.

Since its introduction in 1960, the Marlowe-Crowne Scale has become one of the most frequently used scales in psychological assessment. The reasons for this are twofold. First, the Marlowe-Crowne Scale is the standard measure used to control whether questionnaire responses are biased by desirable responding. This is usually done by demonstrating that the questionnaires under investigation do not correlate with the Marlowe-Crowne Scale or, if they do, that the correlations of these questionnaires with other variables of interest are not significantly attenuated when scores from the Marlowe-Crowne Scale are partialled out (Mummendey, 1981; Paulhus, 1986). Second, the Marlowe-Crowne Scale is the standard measure used to differentiate repressive from low-anxious coping styles (Weinberger, Schwartz, & Davidson, 1979). According to the two-dimen-

sional measurement of repression introduced by these authors, persons with low levels of manifest anxiety and high levels of social desirability are considered "repressors," whereas persons with low levels of manifest anxiety and low levels of social desirability are considered "true low-anxious" (for a review, see Weinberger, 1990).

After 40 years, however, it is questionable if all the items of the Marlowe-Crowne Scale are still up to date. Items like "I am always courteous, even to people who are disagreeable," "There have been times when I felt like rebelling against people in authority even though I knew they were right," and "At times I have really insisted on having things my own way," the latter two reverse-keyed (Crowne & Marlowe, 1960), may reflect social standards of the late 1950s when the Marlowe-Crowne Scale was constructed. But do they still reflect today's standards? Particularly for university students, who represent the great majority of participants in psychological research (Schultz, 1972; Snyder, Tennen, Affleck, & Cheavens, 2000), some contents of the Marlowe-Crowne Scale may be outdated. Therefore, Stöber (1999) decided to construct a new scale in the Marlowe-Crowne style, but with up-to-date contents. Following Crowne and Marlowe's (1960) criteria for item selection, a first sample of students was asked to write down behaviors they considered socially desirable but infrequent or socially undesirable but frequent. After deletion of synonyms and items with psychopathological implications, the remaining answers were used to construct a first pool of 33 items. These items were then given a true/false answer format and, together with the German version of the Marlowe-Crowne Scale (Lück & Timaeus, 1969), presented to a second sample of students. Of the 33 items, 17 items were retained that showed (a) item difficulties between .20 and .80, (b) corrected item-total correlations greater than .20, and (c) significant correlations with the Marlowe-Crowne Scale. Based on the number of items, the resulting scale was named Social Desirability Scale-17 (SDS-17; see Appendix). A third sample was used to gather preliminary data on the reliability and validity of the new scale. With a Cronbach's alpha of .72 and a test-retest correlation of .82 across four weeks, the SDS-17 showed satisfactory reliability. A correlation of .74 with the Marlowe-Crowne Scale demonstrated substantial convergent validity. Moreover, a nonsignificant correlation with trait anxiety as measured with the German version of the State-Trait Anxiety Inventory (Laux, Glanzmann, Schaffner, & Spielberger, 1981) indicated that the SDS-17 may show discriminant validity with respect to neuroticism or negative affectivity (Watson & Clark, 1984). Finally, a fourth sample of students was asked to rate the items of both the SDS-17 and the Marlowe-Crowne Scale for social desirability. Results showed that the SDS-17 items obtained significantly higher ratings than the Marlowe-Crowne items. With an effect size of $d = .64$, the difference between SDS-17 and Marlowe-Crowne items were well above the medium effect size of $d = .50$ that Cohen (1988) considered "large enough to be visible to the naked eye" (p. 26). Thus, for a contemporary measurement of social desirability, the SDS-17 may represent a reliable and valid alternative to the Marlowe-Crowne Scale.

Despite the favorable results of these first investigations, some questions were left open (Stöber, 1999). There are five main points of concern. First, it would be important to demonstrate convergent validity with other contemporary measures of social desirability. Second, it would be important to further demonstrate validity by showing that the SDS-17 is sensitive to social-desirability provoking instructions (Amelang & Bartussek, 1970; Paulhus, 1984). Third, it would be useful to explore the relationship of the SDS-17 with respect to the two components of desirable responding: impression management and self-deception (Paulhus, 1984, 1986). Fourth, whereas Stöber (1999) found that the SDS-17 did not correlate with trait anxiety, it would be important to demonstrate discriminant va-

lidity for other, more comprehensive personality traits. Fifth and finally, it would be important to explore whether the SDS-17 is a reliable and valid measure of social desirability for older adults as well as young adults.

Four studies are presented to provide further information on the convergent validity, discriminant validity, and relationship with age of the SDS-17. With respect to convergent validity, the SDS-17 was compared to the Lie Scale of the revised Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991; Ruch, 1999), the Sets of Four Scale (Borkenau & Ostendorf, 1992), and the Mummendey-Eifler Scale (Mummendey & Eifler, 1993). To examine sensitivity toward social-desirability provoking instructions, the SDS-17 was administered under standard instructions and under job-application instructions, and differences were compared. To examine the relationships with impression management and self-deception, the SDS-17 was compared to the respective scales from the Balanced Inventory of Desirable Responding (Paulhus, 1994). Where discriminant validity is concerned, the SDS-17 was compared to trait scales from two important models of personality, namely Eysenck's model of personality (comprising the traits of neuroticism, extraversion, and psychoticism) and the Five Factor Model of personality (comprising the traits of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness). With respect to the scale's suitability for older adults, the SDS-17 was compared to the Marlowe-Crowne Scale in a sample of participants stratified for gender and age, with age ranging from 18 to 89 years.

Study 1

The aim of Study 1 was to examine the concurrent validity of the SDS-17 with respect to the Lie Scale of the Eysenck Personality Questionnaire, the scale's sensitivity toward social-desirability provoking instructions, and its discriminant validity with respect to neuroticism, extraversion, and psychoticism.

Method

Participants

A sample of $N = 76$ students was recruited at the University of Greifswald. Of these, 61 were female and 14 male (one participant did not indicate his or her gender). Mean age was 22.0 years ($SD = 3.1$; range: = 18-38; two participants did not indicate their age). Participants volunteered in exchange for two hours of extra course credit or a lottery ticket for a chance to win 100 German marks (approximately 50 US dollars).

Measures

In addition to the SDS-17, the short form of the revised Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1991; German version: Ruch, 1999) was administered. The EPQ comprises four scales, three of which capture the superfactors of Eysenck's model of personality: neuroticism (12 items), extraversion (12 items), and psychoticism (14 items). In addition, the EPQ contains a Lie Scale (12 items). All scales are presented with a forced-choice answer format ("Yes" = 1, "No" = 0). The items of the Lie Scale contain questions on socially desirable or undesirable behaviors couched in such extreme terms that acquiescence (or negation, if reverse keyed) can be regarded as "lying"

with the intent to appear more conformist than is actually the case (e.g., "Are all your habits good and desirable ones?" or "Have you ever taken anything [even a pin or button] that belonged to someone else?", reverse keyed). Even though the Lie Scale appears to be very similar to the Marlowe-Crowne Scale in both form and content, there is a difference. Whereas the Marlowe-Crowne Scale was constructed to measure social desirability independent of psychopathology (Crowne & Marlowe, 1960), Eysenck's Lie Scale was constructed to measure dissimulation of psychopathological symptoms (Eysenck & Eysenck, 1964). Accordingly, the Lie Scale has shown substantial negative correlations with psychoticism (e.g., Ruch, 1999).

Procedure

Participants completed a comprehensive two-part battery of questionnaires. In part one, all questionnaires were completed under standard instructions, including the SDS-17 and the EPQ. In part two, the social-desirability provoking instruction from Krampen (1993) was administered: Participants were asked to imagine that they were applying for an important job and that the way they responded to the questionnaires was of great importance for their application. The SDS-17 and the EPQ were then readministered.

Preliminary Analyses

Item analysis indicated that all SDS-17 items showed the same or similar item characteristics as in the previous studies (Stöber, 1999)--with the exception of Item 4, "I have tried illegal drugs (for example, marijuana, cocaine, etc.)." In the previous studies, this item displayed corrected item-total correlations of $r = .27$ and $r = .37$; in the present study, it displayed a corrected item-total correlation of $r = -.07$. It was therefore excluded from the calculation of the SDS-17 total scores. Resulting SDS-17 scores were unrelated to age, $r = .05$, and gender, $t(73) = 0.87$, both *NS*. All scales showed acceptable internal consistencies (Table 1), except for the EPQ psychoticism scale (Cronbach's alpha = .52). To increase the internal consistency of this scale, six items with low item-total correlations were excluded.

Results and Discussion

Results were as expected (Table 1). The SDS-17 showed a substantial correlation with the EPQ-Lie Scale and nonsignificant correlations with the other EPQ scales, indicating high convergent validity with respect to the EPQ measure of social desirability and high discriminant validity with respect to the EPQ measures of neuroticism, extraversion, and psychoticism. In comparison, the EPQ-Lie Scale showed a significant negative correlation with psychoticism. This pattern of results indicates that the SDS-17 and the EPQ-Lie Scale touch somewhat different aspects of social desirability. As intended, the EPQ-Lie Scale seems to capture dissimulation of psychopathological symptoms, whereas the SDS-17, like the Marlowe-Crowne Scale, seems to be a measure of social desirability independent of psychopathology.

Next, the SDS-17's sensitivity to social-desirability provoking instructions was examined. As expected, SDS-17 scores were considerably higher under the imagined job-application instructions than under standard instructions (Table 1). However, the effect of the instructional manipulation was rather unspecific; it not only affected the two social

desirability scales, but all other personality scales, too. This is a well known phenomenon. Neuroticism scales in particular tend to show the same sensitivity to social-desirability provoking instructions as social-desirability scales (e.g., Amelang & Bartussek, 1970). Therefore, a good measure of social desirability should be more sensitive to social-desirability provoking instructions than measures of neuroticism. This was tested with a 2 (scale) \times 2 (instruction) repeated-measures ANOVA on standardized scores from the SDS-17 and the reversed EPQ neuroticism scale. A significant interaction of scale and instruction, $F(1, 75) = 5.41, p < .05$, confirmed that the SDS-17 was more sensitive to the social-desirability provoking instructions than the EPQ neuroticism scale.

Study 2

The aim of Study 2 was to examine the concurrent validity of the SDS-17 with respect to the Sets of Four Scale and its discriminant validity with respect to trait scales from the Five Factor Model of personality: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness.

Method

Participants and Procedure

A sample of $N = 84$ students was recruited at the University of Greifswald. Of these, 69 were female and 12 male (three participants did not indicate their gender). Mean age was 22.1 years ($SD = 3.2$; range = 18-38; three participants did not indicate their age). Participants completed a comprehensive questionnaire battery that included the SDS-17 and the measures described below. Participants volunteered in exchange for two hours of extra course credit or a lottery ticket for a chance to win 100 German marks.

Measures

In addition to the SDS-17, the Sets of Four Scale (Borkenau & Ostendorf, 1992) was administered to measure social desirability. The Sets of Four Scale is based on research by Peabody (1967). To separate evaluative and descriptive aspects in words that describe personality traits, Peabody constructed quadruples of trait-descriptive adjectives, two of which were similar descriptively, but opposite evaluatively and two of which were similar evaluatively, but opposite descriptively. Realizing that this principle could be used to measure social desirability, Borkenau and Ostendorf (1989, 1992) constructed a scale with 20 Peabody quadruples, or "sets of four." Each set of four was combined with two rating scales from -3 to $+3$ (without a neutral point), as depicted in the following example:

lax	-3	-2	-1	+1	+2	+3	firm
severe	-3	-2	-1	+1	+2	+3	lenient

The principle of the Sets of Four is that respondents with a tendency for desirable responding would describe themselves with high ratings for evaluatively positive adjectives (e.g., as both firm and lenient), disregarding the fact that this may be descriptively incon-

sistent. Scores on the Sets of Four Scale are computed by summing across all 40 adjective pairs, with ratings keyed in the evaluatively positive direction (Borkenau & Ostendorf, 1992).

To measure the Big Five personality traits, the NEO Five Factor Inventory (FFI) (Costa & McCrae, 1992; German version: Borkenau & Ostendorf, 1993) was selected. The FFI consists of five scales--neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness--each with 12 items. Answers are given on a five-point rating scale from "strongly disagree" (0) to "strongly agree" (4). FFI scores are calculated by forming mean scores across items.

Preliminary Analyses

As in Study 1, Item 4 ("illegal drugs") showed a corrected item-total correlation close to zero, $r = .07$. Consequently, it was again excluded from the calculation of SDS-17 total scores. Again, SDS-17 scores were unrelated to age, $r = .17$, and gender, $t(79) = 1.21$, both *NS*. All scales showed satisfactory internal consistencies (Table 2).

Results and Discussion

As expected, SDS-17 scores showed a substantial correlation with the scores from the Sets of Four (Table 2). As in Study 1, SDS-17 scores did not correlate significantly with neuroticism or extraversion. In this study, they did not correlate with openness to experience either, but the correlations with agreeableness and conscientiousness were significant. Thus, the results of Study 2 indicate that the SDS-17 has high convergent validity with social desirability as measured with the Sets of Four and high discriminant validity with respect to the FFI measures of neuroticism, extraversion, and openness for experience. The interpretation of the substantial correlations with agreeableness and conscientiousness, however, is a more controversial issue, related to the question of whether social desirability scales measure more substance than style (e.g., McCrae & Costa, 1983). Depending on one's point of view, the correlations may be interpreted as convergent validity or as lack of discriminant validity. On the one hand, it may be argued that agreeableness and conscientiousness are essential parts of social desirability.¹ Accordingly, scores on social desirability scales usually show substantial correlations with self-ratings for agreeableness and conscientiousness (Borkenau & Ostendorf, 1989, 1992; Paulhus, 1994). On the other hand, it may be argued that social desirability scales should capture social desirability *bias* in questionnaire responses (lying, dissimulation, impression management) independent of respondents' levels of agreeableness and conscientiousness. Otherwise, when social desirability scales are used to control for the effects of desirable responding (see Introduction), it may be unclear if correlations with these scales are due to differences in desirable responding or differences in agreeableness and conscientiousness. Either way, the overlap between measures of social desirability and measures of agreeableness and conscientiousness warrants further attention in future research (see General Discussion).

Study 3

The aim of Study 3 was to examine the relationship of the SDS-17 with the Mummendey-Eifler Scale and the scales of the Balanced Inventory of Desirable Responding,

the latter to determine the relationship of the SDS-17 with the two components of social desirability: impression management and self-deception. It was expected that SDS-17 scores would show higher correlations with impression management than with self-deception.

Method

Participants and Procedure

A sample of $N = 101$ students was recruited at the University of Halle. Of these, 79 were female and 20 male (two participants did not indicate their gender). Mean age was 22.2 years ($SD = 3.3$; range = 19-40; four participants did not indicate their age). Respondents completed a comprehensive questionnaire battery that included the SDS-17 and the measures described below. Participants volunteered in exchange for two hours of extra course credit or a lottery ticket for a chance to win 100 German marks.

Measures

In addition to the SDS-17, the Mummendey-Eifler Scale (Mummendey & Eifler, 1993) was administered. The Mummendey-Eifler Scale is a new measure of social desirability consisting of items from the Trier Personality Inventory (TPI; Becker, 1989), and was constructed using the method of instructional variation. A student sample responded to selected items from the TPI subscales on Mental Health, Behavior Control, Autonomy, and Expansiveness, first under standard instructions and then under social-desirability provoking instructions (i.e., to answer all items in such a way as to make a favorable, positive impression). The items with the greatest sensitivity to the social-desirability provoking instruction were selected for inclusion in the new scale. With the removal of redundant items, the scale was reduced to 12 items (e.g., "I am in good physical and mental condition" or "There are times when I cannot stand myself", reverse keyed). As the Mummendey-Eifler Scale is intended for use with variable answer formats depending on the research question (Hans Mummendey, personal e-mail communication, June 4, 1999), the scale was administered with the same answer format as the SDS-17 ("true" = 1, "false" = 0).

The Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1994; German version: Musch, 1999) is based on the findings that items used for the measurement of social desirability form two distinct factors, namely impression management and self-deception (Paulhus, 1984). Impression management refers to the conscious dissimulation of item responses with the aim of making a favorable impression on others. In contrast, self-deception refers to a positive bias in item responses with the aim of protecting positive self-esteem (Paulhus, 1986). The BIDR contains 40 items; 20 items capture impression management (e.g., "When I hear people talking privately, I avoid listening" or "I sometimes drive faster than the speed limit", reverse keyed), and 20 items capture self-deception ("I always know why I like things" or "It would be hard for me to break any of my bad habits", reverse keyed). Items are answered on a seven-point rating scale from "Not true" (1) to "Very true" (7). As only extreme answers indicate socially desirable responding, item ratings are dichotomized (Paulhus, 1994): Ratings of 6 or 7 on positively keyed items and ratings of 1 or 2 on reverse-keyed items are recoded to values of 1; all other

ratings are recoded to values of 0. Scale scores are then computed by summing across items.

Preliminary Analyses

As in the two previous studies, Item 4 ("illegal drugs") showed a corrected item-total correlation near zero, $r = -.04$, and was dropped from the calculation of SDS-17 total scores. Again, SDS-17 scores were unrelated to age, $r = -.08$, and gender, $t(97) = 1.04$, both *NS*. Whereas the SDS-17, the Mummendey-Eifler Scale, and the impression-management scale of the BIDR showed acceptable internal consistencies, the self-deception scale of the BIDR did not (Cronbach's alpha = .59). In order to achieve an acceptable internal consistency, four items with item-total correlations near zero were deleted from the scale.

Results and Discussion

As expected, the SDS-17 scores showed a higher correlation with the impression management scale of the BIDR than with the self-deception scale (Table 3). Moreover, the significant correlation between SDS-17 and self-deception was due only to the overlap between impression management and self-deception. When SDS-17 scores were regressed on impression management and self-deception simultaneously, only impression management showed a significant semi-partial correlation ($sr = .31, p < .01$), whereas the semi-partial correlation of self-deception was nonsignificant ($sr = .08, NS$).² Somewhat unexpectedly, the SDS-17 did not correlate significantly with the Mummendey-Eifler Scale. However, the SDS-17 and the Mummendey-Eifler Scale seem to reflect different components of desirable responding. When scores from the Mummendey-Eifler Scale were regressed on impression management and self-deception simultaneously, only self-deception showed a significant semi-partial correlation ($sr = .37, p < .001$), whereas the semi-partial correlation of impression management was near zero ($sr = .07, NS$). Thus, the SDS-17 relates to the impression-management component of social desirability, whereas the Mummendey-Eifler Scale relates to the self-deception component. Consequently, the nonsignificant correlation of the SDS-17 and the Mummendey-Eifler Scale may indicate discriminant validity, rather than lack of convergent validity.

Study 4

A serious limitation of both Studies 1 to 3 and the previous studies (Stöber, 1999) is that they were all restricted to samples of university students. Moreover, in all samples, male respondents were underrepresented. Thus, the aim of Study 4 was to examine whether the SDS-17 was also suitable for middle-aged and older adults who were not students. Gender effects were also to be reexamined. Responses from a large community sample stratified for gender and age (with age ranging from 18 to 89 years) to the SDS-17 and the Marlowe-Crowne Scale were therefore analyzed. As mentioned in the Introduction, the items of the Marlowe-Crowne Scale may reflect social standards of the late 1950s. Hence, it was expected that scores from the Marlowe-Crowne Scale would show greater increases with age than scores from the SDS-17.

Method

Participants and Procedure

The sample used in Study 4 was part of a larger questionnaire study on selection, optimization, and compensation across the lifespan (Freund & Baltes, 1999) conducted at the Max Planck Institute for Human Development in Berlin. A community sample of $N = 205$ adults was contacted by mail. Participants were asked to fill in a number of questionnaires (including the two social desirability scales) and then to mail back the package. They received 20 German marks (approximately 10 US dollars) for filling in the questionnaires which took about one hour. Of the 205 persons contacted, 181 (88%) sent back completed questionnaires. As two participants returned social desirability scales with substantial missing data, the final sample size for the present analyses was $N = 179$. Mean age was $M = 54.7$ years ($SD = 19.8$; range = 18-89). Table 4 shows the age distribution when broken down into decades. 52% of the participants were female. Gender was approximately balanced across the seven age groups. Only 11% of the participants were students or apprentices. The rest were employed (37%), unemployed (7%), or retired (46%).

Measures

In addition to the SDS-17 (presented without Item 4), the questionnaire package included the Marlowe-Crowne Scale (Crowne & Marlowe, 1960; German version: Lück & Timaeus, 1969) to measure social desirability. The German version of the Marlowe-Crowne Scale (Lück & Timaeus, 1969) consists of 23 items. Like the SDS-17, it contains a forced-choice answer format ("true" = 1, "false" = 0).

Results and Discussion

The internal consistency of the SDS-17 scores was highly satisfactory. Across the total sample, Cronbach's alpha was .80. Moreover, results showed that the internal consistency of the SDS-17 scores did not differ greatly between age groups (Table 4). The convergent validity with the Marlowe-Crowne Scale was also satisfactory. Across the total sample, the SDS-17 scores showed a convergent correlation of $r = .68$, $p < .001$ with scores from the Marlowe-Crowne Scale. However, this correlation was moderated by age, as indicated by a significant interaction with age group, $F(6, 165) = 6.42$, $p < .001$. Whereas the correlation between the two scales was between .75 and .85 in middle-aged participants (30-69 yr), it was no more than .55 in younger participants (18-29 yr) and older participants (70-79 yr), and near zero in the oldest participants (80-89 yr). As both scales showed acceptable internal consistencies in the latter group, the reasons for the non-convergence of SDS-17 and Marlowe-Crowne Scale in the oldest participants remain unclear and demand further research. Until then, researchers should be cautioned against using the SDS-17 (or the Marlowe-Crowne Scale) with participants of 80 years and above.

To investigate whether SDS-17 scores were affected by gender or age, a 2 (gender) \times 7 (age group) between-participants ANOVA was calculated. As in the previous three studies, gender did not have any effect, $F(1, 165) = 0.95$, *NS*. In contrast, age now had a highly significant effect, $F(6, 165) = 5.71$, $p < .001$. Older participants showed significantly higher SDS-17 scores than younger participants. The same effect was found for scores from the Marlowe-Crowne Scale, $F(6, 165) = 13.26$, $p < .001$. Moreover, as expected, the effect was more pronounced for the Marlowe-Crowne Scale than for the SDS-

17: A 2 (scales) \times 7 (age groups) mixed ANOVA on standardized scores yielded a significant interaction effect, $F(6, 172) = 2.15, p < .05$, demonstrating that the scores from the Marlowe-Crowne Scale showed significantly larger increments with age than those from the SDS-17. For the SDS-17, the difference in mean scores between the oldest and the youngest age group was 1.08 standard deviations; for the Marlowe-Crowne Scale, it was 1.75 standard deviations (see Table 4, *Z* scores). To compare these results with the results of the previous three studies, correlations were calculated. Here, the effect was even more pronounced. The correlation between the Marlowe-Crowne Scale and age was $r = .55$, while that between the SDS-17 and age was $r = .38$, both $ps < .001$. The difference between the two correlations was highly significant, $Z = 3.35, p < .001$ (Meng, Rosenthal, & Rubin, 1992). Thus, whereas both scales showed high convergence across all age groups, age (cohort) had a greater influence on the mean scores from the Marlowe-Crowne Scale than on those from the SDS-17.

It is an established finding that older individuals show higher scores in social desirability scales than younger individuals (e.g., Fraboni & Cooper, 1989; Ray, 1988). However, it is unclear whether these differences represent a greater tendency for socially desirable responding in older persons, whether they reflect different norms or value systems, or whether they are simply an expression of higher levels of agreeableness and conscientiousness in older persons. In a recent cross-cultural study with participants aged 18 to 84 years, McCrae et al. (1999) found that participants from older cohorts displayed significantly higher levels of both agreeableness and conscientiousness than participants from younger cohorts. Consequently, in the present study, the higher means of the older cohorts may simply be a reflection of higher agreeableness and conscientiousness, all the more so as Study 2 found substantial correlations between the SDS-17 scores and these two personality dimensions. Therefore, future studies investigating the relationship of social desirability and age should also include measures of agreeableness and conscientiousness in order to control the influence of age differences in these personality variables on social desirability scale scores.

General Discussion

Four studies further examining the validity of the Social Desirability Scale-17 (Stöber, 1999) were presented. Overall, results demonstrated that the SDS-17 has substantial validity. Where convergent validity is concerned, the SDS-17 showed substantial correlations with the Lie Scale of the revised Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991), the Sets of Four Scale (Borkenau & Ostendorf, 1992), and the Marlowe-Crowne Scale (Crowne & Marlowe, 1960). Moreover, the SDS-17 showed high sensitivity toward social-desirability provoking instructions: When participants were instructed to imagine that they were applying for an important job and to respond accordingly, SDS-17 scores were 1.7 standard deviations above those attained under standard instructions. In addition, analyses with the Balanced Inventory of Desirable Responding (Paulhus, 1994) indicated that the SDS-17 relates to the impression-management component of desirable responding, not the self-deception component. Where discriminant validity is concerned, SDS-17 scores showed nonsignificant correlations with neuroticism, extraversion, psychoticism, and openness to experience as measured with the revised Eysenck Personality Questionnaire (Eysenck & Eysenck, 1991; Ruch, 1999) and the NEO Five Factor Inventory (Costa & McCrae, 1993) respectively. With respect to agreeableness and conscientiousness as measured with the Five Factor Inventory, discriminant va-

lidity was lower than would be desirable for a measure designed to capture social-desirability *bias* in self-reports. Finally, it was demonstrated that the use of the SDS-17 is not restricted to young adults. In a large community sample with participants from 18 to 89 years, SDS-17 scores showed satisfactory internal consistency across all age groups. For all but the oldest participants, the SDS-17 scores displayed substantial convergent correlations with the Marlowe-Crowne Scale. At the same time, the present results showed that scores from the SDS-17 were less influenced by participants' age (cohort) than those of the Marlowe-Crowne Scale. Therefore, results from the present studies indicate that the SDS-17 is a reliable and valid measure of social desirability for adults of 18 to 80 years. Together with the findings from the previous studies (Stöber, 1999), they further confirm that the SDS-17 may represent a real alternative for a contemporary measurement of social desirability.

One item of the SDS-17, however, should be excluded from the calculation of total scores. This is Item 4, "I have tried illegal drugs (for example, marijuana, cocaine, etc.)." Whereas this item showed satisfactory item-total correlations in the previous studies (Stöber, 1999), it showed item-total correlations around zero in the present studies (Studies 1 to 3). A potential reason for this discrepancy is that the previous studies were conducted at the Free University of Berlin, which is located in former West Berlin. In contrast, the present studies were conducted at the University of Greifswald and the University of Halle, both of which are located in former East Germany. In East Germany, illegal drugs were not available before German reunification in 1989, and few young East Germans had any experience of these drugs. In West Germany, on the other hand, 26% of young adults between 21 to 26 years claim to have such experience (Nordlohne, Reißig, & Hurrelmann, 1993). A negative response to Item 4 may indicate social desirability only for persons who have experience with illegal drugs. For all other persons (e.g., persons who have never had access to illegal drugs or persons who strictly oppose drugs), such a response does not appear to indicate social desirability. Consequently, Item 4 is not generally suitable for the measurement of socially desirable responding and should be deleted from the SDS-17.³

For future studies on the SDS-17, it may also be useful to gather more information on the stability of SDS-17 scores. Even though the SDS-17 has demonstrated test-retest correlations over .80 across intervals from two to four weeks (Göhner, 1999; Stöber, 1999), it would be useful to have information on the relative contribution to the stability of SDS-17 scores of person variables, situational variables, and their interactions, as may be determined with analyses of latent trait-state models of social desirability (Schmitt & Steyer, 1993). Moreover, to further investigate the convergent validity of the SDS-17, it would be important to gather information from sources other than self-reports, for example, observer ratings from peers or spouses (Borkenau & Ostendorf, 1992; McCrae & Costa, 1983). Moreover, analysis of self-observer agreement on measures of social desirability, agreeableness, and conscientiousness may help to disentangle differences in desirable responding from differences in agreeableness and conscientiousness, and thus allow estimation of the relative contribution of social desirability bias and valid self-descriptions in the relationships between the three personality measures (McCrae, 1994). Finally, it would be helpful to obtain more comprehensive information about age trends in social-desirability scales in order to disclose whether the age-related increases in social desirability reflect a greater tendency for socially desirable responding, higher levels of agreeableness and conscientiousness, or different value systems and norms.

References

- Amelang, M., & Bartussek, D. (1970). Untersuchungen zur Validität einer neuen Lügenskala [Studies on the validity of a new lie-scale]. *Diagnostica, 16*, 103-122.
- Becker, P. (1989). *Der Trierer Persönlichkeitsfragebogen: TPF* [The Trier Personality Inventory: TPI]. Göttingen, Germany: Hogrefe.
- Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality, 3*, 31-45.
- Borkenau, P., & Ostendorf, F. (1992). Social desirability scales as moderator and suppressor variables. *European Journal of Personality, 6*, 199-214.
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI)* [NEO Five Factor Inventory (NEO-FFI)]. Göttingen, Germany: Hogrefe.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*, 90-93.
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York: Dryden.
- Eysenck, H. J., & Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London: University of London Press.
- Eysenck, H. J., & Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult)*. London: Hodder and Stoughton.
- Fraboni, M., & Cooper, D. (1989). Further validation of three short forms of the Marlowe-Crowne Scale of Social Desirability. *Psychological Reports, 65*, 595-600.
- Freund, A. M., & Baltes, P. B. (1999). *Assessing selection, optimization, and compensation: A questionnaire approach*. Manuscript submitted for publication, Max Planck Institute for Human Development, Berlin, Germany.
- Göhner, W. (1999). *Selbständiges Training in der Physiotherapie* [Autonomous training in physiotherapy]. Unpublished diploma thesis, University of Trier, Germany.
- Krampen, G. (1993). Effekte von Bewerbungsinstruktionen und Subskalenextraktion in der Fragebogendiagnostik [Effects of job-application instructions and subscale extraction in assessment via questionnaires]. *Diagnostica, 39*, 97-108.
- Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C. D. (1981). *STAI: Das State-Trait-Angstinventar [STAI: The State-Trait Anxiety Inventory]*. Weinheim, Germany: Beltz Testgesellschaft.
- Lück, H. E., & Timaeus, E. (1969). Skalen zur Messung Manifester Angst (MAS) und sozialer Wünschbarkeit (SDS-E und SDS-CM) [Scales for the measurement of manifest anxiety (MAS) and social desirability (SDS-E and SDS-CM)]. *Diagnostica, 15*, 134-141.
- McCrae, R. R. (1994). The counterpoint of personality assessment: Self-reports and observer ratings. *Assessment, 1*, 159-172.
- McCrae, R. R., & Costa, P. T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*, 882-888.

- McCrae, R. R., Costa, P. T., Jr., de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., Marušić, I., Bratko, D., Caprara, G. V., Barbaranelli, C., Chae, J.-H., & Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology, 35*, 466-477.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172-175.
- Mummendey, H. D. (1981). Methoden und Probleme der Kontrolle sozialer Erwünschtheit (Social Desirability) [Methods and problems in the control of social desirability]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 2*, 199-218.
- Mummendey, H. D., & Eifler, S. (1993). *Eine neue Skala zur Messung Sozialer Erwünschtheit* [A new scale for the measurement of social desirability] (Bielefelder Arbeiten zur Sozialpsychologie Nr. 167). Bielefeld, Germany: University of Bielefeld.
- Musch, J. (1999). *German version of the Balanced Inventory of Desirable Responding (BIDR version 6)*. Unpublished manuscript, University of Bonn, Germany.
- Nordlohne, E., Reißig, M., & Hurrelmann, K. (1993). Drogenmißbrauch in Ost und West: Zur Situation des Drogengebrauchs bei Jugendlichen in den alten und neuen Ländern der Bundesrepublik [Drug abuse in East and West: On the use of illegal drugs by young people in the old and new states of the Federal Republic of Germany]. *Sucht, 39* (1), 10-34.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaires* (pp. 143-165). Berlin, Germany: Springer.
- Paulhus, D. L. (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR version 6*. Unpublished manuscript, University of British Columbia, Vancouver, Canada.
- Peabody, D. (1967). Trait inferences: Evaluative and descriptive aspects. *Journal of Personality and Social Psychology, 7* (4, PT. 2), 1-18.
- Ray, J. J. (1988). Lie scales and the elderly. *Personality and Individual Differences, 9*, 417-418.
- Ruch, W. (1999). Die revidierte Fassung des Eysenck Personality Questionnaire und die Konstruktion des deutschen EPQ-R bzw. EPQ-RK [The revised version of the Eysenck Personality Questionnaire and the construction of the German EPQ-R and EPQ-RK]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 20*, 1-24.
- Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. *Personality and Individual Differences, 14*, 519-529.
- Schultz, D. P. (1972). The human subject in psychological research. In C. L. Sheridan (Ed.), *Readings for experimental psychology* (pp. 263-282). New York: Holt.
- Snyder, C. R., Tennen, H., Affleck, G., & Cheavens, J. (2000). Social, personality, clinical, and health psychology tributaries: The merging of a scholarly "river of dreams." *Personality and Social Psychology Review, 4*, 16-29.
- Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität [The Social Desirability Scale-17 (SDS-17): Development and first results on reliability and validity]. *Diagnostica, 45*, 173-177.
- Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper and Row.
- Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96*, 465-490.

- Weinberger, D. A. (1990). The construct validity of the repressive coping style. In J. L. Singer (Ed.), *Repression and dissociation: Implications for personality, psychopathology, and health* (pp. 337-386). Chicago, IL: University of Chicago Press.
- Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral and physiological responses to stress. *Journal of Abnormal Psychology, 88*, 369-380.

Author Note

Part of this article was presented at the fifth meeting of the Fachgruppe Differentielle Psychologie, Persönlichkeitspsychologie und Psychologische Diagnostik, Wuppertal, Germany, October 1999.

I would like to express my gratitude to Paul Baltes, Alexandra Freund, Shu-Chen Li, and Ulman Lindenberger for generously providing me with the data used in Study 4. Moreover, I would like to thank Fritz Ostendorf and Jochen Musch for providing me with the Sets of Four Scale and the German version of the BIDR, respectively. I also would like to thank Birgit Andrich, Thilo Deckersbach, Alexandra Freund, Susannah Goss, Stefan Hofmann, Joachim Krueger, Monisha Pasupathi, Ulrich Schimmack, and Barbara Zebb for translations, back-translations, valuable comments, and helpful discussions associated with the English version of the SDS-17 (Appendix). Finally, many thanks go to Claudia Dalbert, Jutta Joormann, Thomas Rodenhausen, Hubert Sydow, and an anonymous reviewer for helpful comments and suggestions on earlier versions of this article.

Footnotes

¹I would like to thank Hubert Sydow and Hannelore Weber for pointing out this interpretation.

²In standard multiple regression, the squared semi-partial correlation represents the unique contribution that a variable makes in the prediction of a dependent variable (Tabachnick & Fidell, 1989, pp. 151-153).

³Nevertheless, across Studies 1 to 3, the correlation between the original version (with Item 4) and the revised version (without Item 4) of the SDS-17 was $r = .99$. Consequently, all results attained with the revised SDS-17 should generalize to those yielded by the original scale and vice versa.

Table 1

The SDS-17 and the Scales of the Eysenck Personality Questionnaire under Standard and Social-Desirability Provoking Instructions

Scale	Instruction										
	Standard				SD provoking ^b						
	<i>M</i>	<i>SD</i>	α	Correlation				<i>M</i>	<i>SD</i>	<i>t</i> (75)	
			1	2	3	4					
1. SDS-17	Social desirability	7.19	3.35	.74	--				12.92	3.34	14.23***
2. EPQ-Lie	Social desirability	1.47	1.60	.61	.60***	--			4.80	3.27	9.65***
3. EPQ-N	Neuroticism	4.99	2.54	.69	-.09	-.17	--		2.11	2.13	-10.68***
4. EPQ-E	Extraversion	7.56	2.83	.78	-.07	-.18	.17	--	8.66	2.73	4.83***
5. EPQ-P ^a	Psychoticism	2.11	1.66	.60	-.16	-.33**	.16	.07	0.86	1.41	-7.91***

Note. $N = 76$. SDS-17 = Social Desirability Scale-17, without Item 4 ("illegal drugs"). EPQ = short form of the revised Eysenck Personality Questionnaire. α = Cronbach's alpha. $t(75)$ = t value of mean difference.

^aScale reduced to eight items (see text for details). ^bImagined job-application instruction (Krampen, 1993).

** $p < .01$. *** $p < .001$.

Table 2

The SDS-17, the Sets of Four Scale, and the Scales of the NEO Five Factor Inventory

Scale	<i>M</i>	<i>SD</i>	α	Correlation						
				1	2	3	4	5	6	
1. SDS-17	7.11	3.40	.75	--						
2. Sets of Four	39.35	13.19	.82	.52***	--					
3. FFI-N	1.83	0.55	.78	-.14	-.49***	--				
4. FFI-E	2.47	0.53	.80	.01	.04	-.12	--			
5. FFI-O	2.88	0.49	.75	.03	.03	.13	.14	--		
6. FFI-A	2.70	0.45	.70	.34**	.55***	-.32**	.09	.03	--	
7. FFI-C	2.61	0.50	.79	.38***	.34**	-.10	.00	-.20	.22*	--

Note. $N = 84$. SDS-17 = Social Desirability Scale-17, without Item 4 ("illegal drugs"). FFI = NEO Five Factor Inventory. α = Cronbach's alpha.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

The SDS-17, the Mummendey-Eifler Scale, and the Scales of the Balanced Inventory of Desirable Responding

Scale		<i>M</i>	<i>SD</i>	α	Correlation		
					1	2	3
1. SDS-17	Social desirability	8.53	3.45	.75	--		
2. MES	Social desirability	7.92	2.62	.74	.17	--	
3. BIDR-IM	Impression management	4.99	3.10	.68	.37***	.22*	--
4. BIDR-SDE ^a	Self-deception	3.92	2.64	.65	.21*	.43***	.38***

Note. $N = 101$. SDS-17 = Social Desirability Scale-17, without Item 4 ("illegal drugs"). MES = Mummendey-Eifler Scale. BIDR = Balanced Inventory of Desirable Responding. α = Cronbach's alpha.

^aScale reduced to 16 items (see text for details).

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

Table 4
The SDS-17 and the Marlowe-Crowne Scale Across Age Groups of 18-89 Years

Group	Age	n	Scale								r
			SDS-17				MCS				
			M	SD	α	Z	M	SD	α	Z	
1	18-29	25	9.32	3.54	.78	-0.58	10.56	3.75	.68	-0.89	.51**
2	30-39	26	10.08	3.49	.80	-0.36	12.54	4.40	.78	-0.44	.85***
3	40-49	25	10.03	2.90	.70	-0.37	13.25	3.51	.66	-0.28	.78***
4	50-59	22	11.21	2.91	.73	-0.02	13.81	4.44	.79	-0.16	.79***
5	60-69	30	13.19	2.08	.61	0.57	16.52	3.72	.74	0.45	.75***
6	70-79	27	11.83	3.30	.80	0.16	16.20	3.72	.75	0.38	.55**
7	80-89	24	12.95	3.29	.84	0.50	18.34	2.67	.62	0.86	-.05

Note. N = 179. SDS-17 = Social Desirability Scale-17, without Item 4 ("illegal drugs"). MCS = Marlowe-Crowne Scale. α = Cronbach's alpha. Z = standardized scores. r = correlation between SDS-17 and MCS.

** $p < .01$. *** $p < .001$.

Appendix: The Social Desirability Scale-17 (SDS-17)

Instruction

Below you will find a list of statements. Please read each statement carefully and decide if that statement describes you or not. If it describes you, check the word "true"; if not, check the word "false".

Items

1. I sometimes litter.
2. I always admit my mistakes openly and face the potential negative consequences.
3. In traffic I am always polite and considerate of others.
4. I have tried illegal drugs (for example, marijuana, cocaine, etc.).
5. I always accept others' opinions, even when they don't agree with my own.
6. I take out my bad moods on others now and then.
7. There has been an occasion when I took advantage of someone else.
8. In conversations I always listen attentively and let others finish their sentences.
9. I never hesitate to help someone in case of emergency.
10. When I have made a promise, I keep it--no ifs, ands or buts.
11. I occasionally speak badly of others behind their back.
12. I would never live off other people.
13. I always stay friendly and courteous with other people, even when I am stressed out.
14. During arguments I always stay objective and matter-of-fact.
15. There has been at least one occasion when I failed to return an item that I borrowed.
16. I always eat a healthy diet.
17. Sometimes I only help because I expect something in return.

Note

Answer categories are "true" (1) and "false" (0). Items 1, 4, 6, 7, 11, 15, and 17 are reverse keyed. Item 4 was deleted from the final version of the SDS-17.