

Measurement and Control of Response Bias

Delroy L. Paulhus

A *response bias* is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure). For example, a respondent might choose the option that is most extreme or most socially desirable. A response bias might be a *response set*, that is, a temporary reaction to a situational demand, for example, time pressure or expected public disclosure. Alternatively, a bias may be induced by context effects such as the item format or the nature of previous items in the questionnaire (for a review, see Tourangeau & Rasinski, 1988). To the extent that an individual displays the bias consistently across time and situations, the bias is said to be a *response style* (Jackson & Messick, 1958; Wiggins, 1973).

Response biases continue to be a disturbing issue in psychological assessment, particularly with self-report measures such as those in the present volume. People's reports of their own traits, attitudes, and behavior may involve systematic biases that obscure measurement of content variables. For example, there is evidence that standard self-report methodologies distort the reporting of racist attitudes (Sigall & Page, 1971), nonnormative sexual attitudes (Knudson, Pope, & Irish, 1967), desirable behaviors (Phillips & Clancy, 1972), deviant behaviors (Clark & Tift, 1966), and abortion (Wiseman, 1972). Another repercussion of response bias is that an assessment instrument may confound content with style. That is, the instrument may simultaneously assess the individual's response style as well as the content variable. If so, every observed correlation with this instrument will be open to (at least) two possible explanations.

Among the response biases cited in the literature are deviant responding (Berg, 1967), careless responding (Meehl & Hathaway, 1946), consistent responding (Dillehay & Jernigan, 1970), and omitting items (Cronbach, 1946). In this chapter, however, attention will be restricted to the three most prominent response biases: (1) socially desirable responding (including lying, faking bad, etc.), (2) acquiescence (tendency to agree), and (3) extremity bias (tendency to use extreme ratings).

Socially Desirable Responding

The most frequently studied response bias is socially desirable responding (SDR), the tendency to give answers that make the respondent look good. Over 50 years ago,

psychometricians had already raised the issue of SDR effects on the validity of questionnaires (e.g., Bernreuter, 1933; Vernon, 1934). Ten years later, Meehl and Hathaway (1946) were able to cite eight measures specifically developed to index SDR in self-reports.¹ Since the 1950s, SDR has been a prominent concern in measuring personality (e.g., Edwards, 1953), psychopathology (e.g., Gough, 1947; McKinley, Hathaway, & Meehl, 1948), attitudes (Lanski & Leggett, 1960), and self-reports of sensitive behavior (Goode & Hart, 1952).

The first section below reviews methods developed to control SDR in each of these realms. Where SDR cannot be controlled, it might still be measured: Accordingly, the second section below reviews the best available measures of SDR.

Controlling the Influence of SDR

Available methods for controlling SDR are varied and often complex [for reviews, see DeMaio (1984), Nederhof (1985), or Paulhus (1981)]. Only a brief summary can be rendered here. Four types of methods may be distinguished: rational, factor analytic, covariate, and demand reduction methods.

1. Rational techniques are control features that are built into the self-report instrument. Their purpose is to prevent the subject from responding in a socially desirable fashion. For example, one may use a forced-choice format in which the two statements are equated for social desirability. If single statements are used, they might be restricted to those that are roughly neutral with respect to social desirability. Alternatively, one could select statements with high content saturation, that is, statements in which the relative influence of content over desirability is high (see Jackson, 1967).

2. Factor analytic techniques may be applied during test construction if the procedure involves choosing the highest loading items in a factor analysis. If one principal component appears to represent SDR, it may be deleted before the factors are rotated (Paulhus, 1981). Even better, if a measure of SDR is administered along with the content items, then the first factor may be rotated to the SDR measure (e.g., Morf & Jackson, 1972; Norman, 1969). In both techniques the highest loading items on the remaining factors necessarily tap individual differences above and beyond the effects of SDR. Hence, these items may be used to construct SDR-free content measures.

3. In covariate techniques, no attempt is made to prevent respondents from answering in a desirable fashion. Instead, a measure of SDR, for example, one of the measures provided at the end of this chapter, is administered along with the content measures. SDR may then be partialled out of correlations between two content scales to control for spurious correlation.

Alternatively, to improve the validity of individual scores, the raw score may be adjusted by an amount commensurate with the contamination by SDR. This adjustment may be effected by regressing the content score on SDR: The residual then represents the content score corrected for SDR (e.g., Norman, 1967). This procedure is systematized in scoring the MMPI, in which certain clinical scale scores are adjusted using the K scale as a measure of SDR (McKinley *et al.*, 1948). A similar procedure is now used in adjusting several scales of the 16PF (Karson & O'Dell, 1976).

4. A wide-ranging form of control, demand reduction, includes those methods that

¹Precedence should be granted to Hartshorne and May (1928), who in the twenties had already developed a lie scale to directly assess dishonesty in school children.

reduce the situational press for desirable responding. Perhaps the most obvious strategy is to assure respondents of anonymity. In classroom administrations, perceived anonymity is increased by (1) physically separating respondents (especially acquaintances), (2) insisting that they put no identifying marks on the questionnaire (Becker, 1976), and (3) telling them beforehand that they will seal the completed questionnaire in an envelope and drop it in a box on the way out. Presumably because of perceived anonymity, mail surveys appear to be less susceptible to SDR (Wiseman, 1972). Computerized assessments show lower SDR than face-to-face interviews, but higher SDR than paper-and-pencil tests (Martin & Nagao, 1989; Davis & Cowles, 1989; Lautenschlager & Flaherty, in press).

If it is necessary to match responses across two administrations, respondents could be asked to give their birthdates, or to use a consistent pseudonym for all administrations. Despite all such precautions, at least some subjects may still suspect that their identity can be determined by the experimenters.

A more aggressive technique, the bogus pipeline (Jones & Sigall, 1971), is essentially a pseudo lie detector. Respondents are hooked up to electronic equipment that the operator claims can assess their attitudes directly through physiological measures. The equipment is said to be, in effect, "a pipeline to the soul." As a putative test of their own self-insight, subjects are asked to guess the machine's reading for each attitude question. The rationale is that subjects want to avoid being embarrassed by the machine: Hence, their guesses should be less contaminated with SDR than are ordinary self-reports.

The efficacy of the technique is documented by its ability to increase admissions of such undesirable behavior as (a) racist attitudes (Sigall & Page, 1971), (b) sexist attitudes (Faranda, Kaminski, & Giza, 1979), (c) inconsistent attitudes (Paulhus, 1982), (d) dislike for a handicapped confederate (Sigall & Page, 1972), and (e) having prior knowledge of test answers (Quigley-Fernandez & Tedeschi, 1978).

Some related techniques are milder versions of the bogus pipeline. For example, before respondents complete a questionnaire they are warned that the measure contains methods for detecting faking. In a complex variant of this technique, respondents are given an especially transparent lie scale, which is then scored, and they are advised whether their score indicates faking (Montag & Comrey, 1982).

A technique useful for face-to-face interviews is the Randomized Response Method (Greenberg, Abdula, Simmons, & Horvitz, 1969; Warner, 1965). The desirability-loaded question (e.g., Have you ever had an abortion?) is posed along with an innocuous question (Are your mother's eyes blue?). The respondent is instructed to flip a coin and (without telling the interviewer how the flip came out) answer question 1 if heads; question 2, if tails. Since the interviewer is not aware which question is being answered, the respondent is under less pressure to respond desirably. Nonetheless, the abortion rate in the sample can ultimately be estimated from three group statistics: the observed proportion of "yes" responses, the assumption that .50 of the respondents answered the abortion question, and the known rate of blue eyes in the population. The bulk of the evidence suggests that the randomized response technique does increase reports of sensitive behaviors, although not quite up to their true rates (Dawes & Smith, 1985).

Psychological assessment through biographical data is a unique approach that tends to reduce SDR demand (Shaffer, Saunders, & Owens, 1986). The rationale is that, when reporting on concrete, verifiable facts, subjects are less tempted to dissimulate.

Finally, some researchers have tried to reduce SDR by employing proxy subjects: Instead of the target person, a close acquaintance is questioned about the target's behavior. According to evidence collected by Sudman and Bradburn (1974), this technique is satisfactory for measuring publicly observable types of behavior, but not for attitudes. The

evidence for its efficacy in measuring personality is mixed (Kane & Lawler, 1978; McCrac, 1982).

5. The last category, stress minimization, refers to basic guidelines for reducing tension during the test administration. Among the factors known to increase desirable responding are: (a) speed instructions (Sutherland & Spilka, 1964), (b) emotional arousal (Paulhus & Levitt, 1987), and distraction (Paulhus, Graf, & Van Selst, 1989). Hence, these factors should all be minimized.

Measurement of SDR

There are a number of reasons why a researcher might want a direct measure of SDR such as those included in this chapter. The most common usage is for supporting the discriminant validity of a content instrument. To ensure that the content instrument is not confounded with SDR, the researcher would administer both measures to the same sample, hoping to see a low intercorrelation. An SDR measure would also be necessary to use the covariate and target rotation techniques described above.

In some cases, rather than correcting a biased score, one might prefer to discard an individual's data if an accompanying SDR measure detected a high degree of SDR. Optimal cutoff scores may be derived for detecting faking good and faking bad (e.g., Helmes & Holden, 1986; Karson & O'Dell, 1976; Lanning, 1989).

Such cutoff scores may then be useful in evaluating the fakeability of content instruments. A content scale is resistant to faking if, under fake-good instructions, the SDR scale exceeds the cutoff point but the content scale does not change. Note that demonstrating that one can fake good on an instrument does not prove that it is confounded with SDR under ordinary conditions (Furnham, 1986). Moreover, faking good can be idiosyncratic (Lautenschlager, 1986).

Situational demands can also be assessed with SDR measures. For example, organizational researchers have used such measures to determine the extent to which various work environments encourage SDR (see Zerbe & Paulhus, 1987, for a review). SDR measures may also assist in determining the best interviewer characteristics (Weiss, 1968) or test conditions (e.g., Martin & Nagao, 1989; Paulhus, 1984; Schriesheim, 1979) to promote disclosure. Computerized testing, for example, has been evaluated with several measures of SDR (Davis & Cowles, 1989; Lautenschlager & Flaherty, in press).

In practice, the primary use for SDR measures has been to assess consistent individual differences, that is, response styles. There is a longstanding concern that response styles interfere with accurate assessment of content variables (e.g., Edwards, 1953; Goode & Hart, 1952; Gough, 1947; McKinley *et al.*, 1948). Such styles have also been claimed to reflect deeper psychological constructs of interest in their own right (Block, 1965; Damarin & Messick, 1965; Sweetland & Quay, 1953). The classic example is Crowne and Marlowe's work (1964) on the need for approval construct, which emerged from studies on their 1960 measure of SDR (Crowne & Marlowe, 1960). Other personality constructs postulated to underlie response styles include repression-sensitization (Byrne, 1964), censure avoidance (Allaman, Joyce & Crandall, 1972; Millham & Jacobson, 1978), and self-deception (Paulhus, 1984; Sackeim & Gur, 1978). Measures of several of these constructs are included at the end of this chapter.

The widespread concern about SDR in test responses is reflected in the fact that major personality batteries invariably include an SDR measure. The MMPI includes two such scales: the Lie scale to detect blatant dissimulation and the K scale to tap more subtle

distortions (McKinley *et al.*, 1948). The Eysenck Personality Inventory (Eysenck & Eysenck, 1964) and the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) both contain a rationally developed lie scale. Comrey (1980) includes the Response Bias scale in the Comrey Personality Scales. The Personality Research Form (Jackson, 1967) contains the Desirability scale assembled from diverse items of extreme desirability. The Differential Personality Questionnaire (Tellegen, 1982) contains the Unlikely Virtues scale as well as a Desirability Inconsistency scale. The California Psychological Inventory contains the Good Impression and Well-Being scales (Gough, 1987). Measures of both faking good and faking bad have been developed for the 16-PF (Winder, O'Dell, & Karson, 1975).

With the exception of the MMPI validity scales, little direct research has been conducted on these commercially published SDR measures. One reason is that these instruments are often tailored to be similar in format and content to the inventory as a whole and hence may have little application outside the inventory. Because of the limitations in empirical evidence, such measures are not presented in detail here; rather, the reader is referred to the manuals cited above for each of the inventories.

Varieties of SDR

A disturbing feature of SDR measures is the low intercorrelation among several of the more well-known instruments, for example, Edwards's SD scale, the Marlowe-Crowne scale and Wiggins's Sd scale. The frustration of nonspecialists is epitomized in the title of one article: "Will the real social desirability scale please stand up?" (Strosahl, Linehan, & Chiles, 1984). Factor analyses of SDR instruments have consistently revealed two primary factors (Borkenau & Ostendorf, 1989; Edwards & Walsh, 1964; Jackson & Messick, 1962; Paulhus, 1984; Wiggins, 1964). One cluster is associated with *Alpha*, the general anxiety factor of the MMPI (Block, 1965). The second cluster is associated with the another MMPI factor called *Gamma* (Wiggins, 1964), which is linked to agreeableness and traditionalism. Paulhus (1984, 1986) provided evidence that these two SDR factors represent (a) self-deceptive positivity (an honest but overly positive self-presentation) and (b) impression management (self-presentation tailored to an audience). This distinction was outlined many years before by Damarin and Messick (1965). As depicted in Fig. 1, the gamut of SDR measures may usefully be characterized in terms of their relative weighting of self-deception and impression management. The Desirability scale of the PRF and Edwards's SD scale load primarily on the self-deception factor; Wiggins's Sd scale and Eysenck's Lie scale load primarily on the impression management factor. The Marlowe-Crowne scale loads on both factors, although more so on impression management.

The term "impression management" was chosen to represent one traditional view of SDR: that some subjects are purposefully tailoring their answers to create the most positive social image. Of the many impressions that one may try to present, this factor represents only one: a socially conventional, dependable persona. The label "impression management" is preferable to "lying," which is an overly harsh and sweeping indictment. After all, such individuals may misrepresent themselves only to avoid social disapproval (Crowne, 1979). Whatever the label,² this tendency will vary according to situational

²This factor has also been given such diverse interpretations as moralistic hypocrisy (Cattell, Pierson, & Finkbeiner, 1976), interpersonal sensitivity (Holden & Fekken, 1989), defensiveness (Weinberger, Schwartz, & Davidson, 1979), extraverted adjustment (McCrae & Costa, 1983), and test-taking intelligence (P. Borkenau, personal communication).

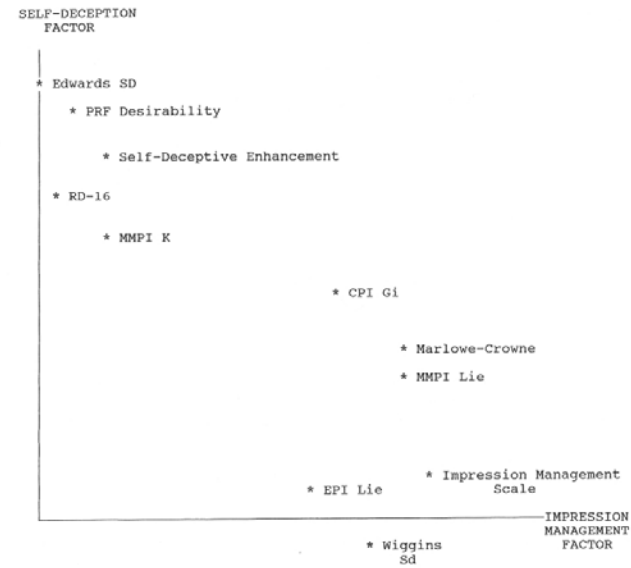


Fig. 1 Typical factor loadings of SDR measures.

demands and transient motives and that variation may obscure the validity of the respondent's self-reports.

The concept of self-deceptive positivity, on the other hand, appears to be intrinsically linked to such personality constructs as adjustment (Taylor & Brown, 1988), optimism (Scheier & Carver, 1985), self-esteem (M. Rosenberg, personal communication), and a sense of general capability (Holden & Fekken, 1989). These conceptual links are supported by significant correlations between measures of self-deception and measures of mental health (Linden, Paulhus, & Dobson, 1986; Sackeim & Gur, 1979), self-esteem (Paulhus & Reid, in press; Winters & Neale, 1985), and various cognitive biases (Paulhus, 1988; Paulhus & Reid, in press). The label "self-deceptive" was chosen in reference to the verifiable distortion by high scorers on certain forms of self-information (Paulhus, 1988).

This two-factor distinction helps resolve several issues in the SDR literature. Research typically shows a positive relation between SDR and adjustment, whereas traditional theories (as well as common sense) presume a negative relation. Self-deceptive SDR bears a strong positive relation with adjustment whereas impression management bears little relation. Thus the relation with adjustment depends on the type of SDR.

The second issue clarified by the self-deception versus impression management distinction is what should be done about the SDR that pervades such central personality variables as perceived control, social dominance, and adjustment. Norman (1967, 1990) argued that removing a general index of SDR from personality ratings clarifies the content dimensions. However, most relevant studies have shown that controlling SDR, if anything, actually *reduces* the predictive validity of content measures (Borkenau & Amelang, 1985; Kozma & Stones, 1988; McCrae, Costa, Dahlstrom, Barefoot, Siegler, & Williams, 1989; McCrae & Costa, 1983; Ruch & Ruch, 1967). Indeed, it now appears that controlling any SDR measure that taps self-deception (e.g., the Edwards scale, the Borkenau-Ostendorf SD scale, the K scale, and to a lesser degree, the Marlowe-Crowne scale) will lower the predictive validity of measures involving a self-deceptive positivity (e.g., anxiety, achievement motivation, dominance, well-being, perceived control, self-esteem). This form of SDR is inextricably linked to content variance and should not be controlled.

In contrast, impression management should be controlled under certain circumstances (Paulhus, 1986), namely, when impression management is conceptually independent of the trait being assessed but still contributes to the self-report scores of that trait. In personnel selection, for example, transient motives ("I'll say anything to get this job because my business just burned down") satisfy those criteria. Many of the techniques listed earlier were designed to control such conscious impression management.

The ideal procedure would involve establishing the distribution of impression management scores in the specific job-selection situation. In addition, the distribution of subjects asked to fake good would be established. If the scores of faking subjects sat several standard deviations above the mean of the no-fake group, then an efficient cutoff point would be easy to specify.

Finally, a set of measures with uncertain status must be mentioned—the so-called honesty or integrity scales. These are questionnaires administered in industrial-organizational settings to determine whether employees or job candidates can be trusted. They fall into two groups. One type contains direct questions about the respondent's integrity, of the form: "Are you now, or have you ever been, a crook?" The second type targets more general traits, such as conscientiousness and impulsiveness, that bear on employee trustworthiness (e.g., Hogan & Hogan, 1989).

Although widely used in business, the direct measures have had a poor reputation among psychologists (e.g., Sackett & Harris, 1984). This reputation may be due in part to the difficulty in obtaining the measures for research purposes. In addition, some honesty scales show positive correlations with lie scales. Thus the same individuals could be labeled honest or dishonest depending on whether honesty scales or lie scales are used. Nonetheless, the more recent evidence about the validity of certain honesty scales is more encouraging (Cunningham, 1989; Sackett, Burris, & Callahan, 1989).

Guidelines

The comments above suggest several practical implications for researchers. Caution must be exercised in interpreting reports of high correlations between self-report instruments and SDR: Such relations must not blithely be assumed to represent contamination and, therefore, deficits in the instruments. Indeed, the SDR component may be a legitimate aspect of the construct being measured.

If individual differences in SDR are evidenced in one's data, one must carefully consider why this has occurred. In a personnel selection situation, for example, a correlation between self-reported motivation and a measure of impression management has a number of plausible interpretations. A stylistic interpretation would suggest that chronic

impression managers are faking high motivation. Alternatively, the nature of the position (e.g., public relations) may be such that chronic impression managers would continue to be motivated and are, therefore, ideal candidates. A third interpretation is that the observed relation is due, not to stylistic impression management, but to a temporary response set: By chance, some respondents are temporarily motivated and are, therefore, presenting a good impression that is unlikely to predict future behavior.

Measures Reviewed Here

In selecting an SDR measure from the eight provided below, the researcher must consider which form of SDR is relevant and select an appropriate measure. The eight measures are

1. Edwards (1957) Social Desirability Scale,
2. Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960),
3. MMPI Lie Scale (Hathaway & McKinley, 1951),
4. MMPI K Scale (Meehl & Hathaway, 1946),
5. Balanced Inventory of Desirable Responding (Paulhus, 1984, 1988),
6. RD-16 (Schuessler, Hittle, & Cardascia, 1978), and
7. Children's Social Desirability Scale (Crandall, Crandall, & Katkovsky, 1965).

It is difficult to order the measures in terms of their value because they each have specialized applications. Instead, they will be presented in descending order of the amount of research published on the measure. No implication should be drawn that SDR measures not included here are deficient. There is simply less information available on other measures (except for the Eysenck Lie scale, which is reviewed by Furnham [1986]).³

The Edwards Social Desirability (SD) scale contains 39 items from the MMPI that have extremely high or low desirability ratings. The scale falls squarely on the first factor of SDR: hence it correlates highly with many standard measures of adjustment (e.g., anxiety, self-esteem, depression) and personality (e.g., perceived control, assertiveness). A high correlation with the SD does not invalidate an individual difference measure but suggests that the measure may involve some positivist bias.

The Marlowe-Crowne (MC) scale contains 33 True-False items about behaviors that are desirable but rare or undesirable but common. The behaviors concern everyday events, not psychopathology. The scale loads on both factors of SDR but more highly on the impression management factor. The MC scale generally correlates less highly than the SD with measures of adjustment.

The MMPI Lie scale contains 15 statements about attitudes and practices that are socially undesirable, but common. Saying "false" to 8 or more is considered evidence that the respondent is faking good on the MMPI. The measure falls toward the second factor of SDR, highly correlated with other lie scales and the Marlowe-Crowne scale. The scale can detect faking good only in naive test-takers.

The MMPI K scale contains 30 items designed to identify abnormal persons whose MMPI scores appear normal. In contrast to the MMPI Lie scale, it is a more subtle

³Nor is there sufficient information concerning two new SDR measures: (1) the Borkenau-Ostendorf Social Desirability scale (Borkenau & Ostendorf, 1989), and (2) the Self-Presentation Scale (Roth, Harris, & Snyder, 1988). A promising technique that requires standardization is Phillips and Clancy's (1972) overclaiming index.

measure of SDR. Its status is rather complex, given that it is said to index psychological health in normal samples, but defensiveness in maladjusted samples.

The CPI Good Impression scale contains 40 items designed to measure faking good. The high scorer wants to present the impression of being dependable, cooperative, and moral. A good deal of validity data are available for this measure (Gough, 1987; Tellegen, 1982).

The Balanced Inventory of Desirable Responding (BIDR) contains separate measures of impression management (audience-driven self-presentation) and self-deceptive enhancement (an honest positivistic bias). The sum of the two measures correlates highly with the MC scale.

The RD-16 instrument was specifically designed to detect SDR in surveys of attitudes and opinions on the general population. An impressive set of norms is available from a national probability sample. The items were screened to preclude differences across such subgroups as race and education.

The Children's Social Desirability Scale was designed for children from grades 6–12. The item content largely follows that of the Marlowe–Crowne scale but is worded in children's language. Additional items involve child-specific content.

Edwards Social Desirability Scale (SD)

(Edwards, 1957)

Variable

Edwards (1957) described the SD as measuring "the tendency to give socially desirable responses in self-description" (p. 35), more specifically, an individual's characteristic level of self-presentation without special instructions or motivation to do so (p. 230).

Description

Edwards (1957) asked ten judges to rate whether "True" or "False" was the most desirable response to 79 items assembled from the K, F, and Lie scales of the MMPI. The 39 items on which the judges unanimously agreed formed the SD. Most of the items (30 of 39) are keyed negatively. As on the MMPI, respondents must answer "True" or "False," with one point added for each response that matches the key. Hence, possible scores range from 0 to 39, higher scores indicating more socially desirable responding. Given their source, it is not surprising that item content is heavily laden with references to psychological distress.

Samples

Edwards (1957) reported means of 28.6 (s.d. = 6.5) and 27.1 (s.d. = 6.5) for males and females in a sample of 192 college students. Edwards and Walsh (1964) found a mean of 28.8 in a sample of 130 paid students. More recently, Paulhus (1984) reported means of 28.4 (s.d. = 5.5) and 30.3 (s.d. = 4.9) for students in anonymous ($n = 60$) and public disclosure conditions ($n = 40$), respectively. In a sample of 503 students, Tanaka–Matsumi and Kameoka (1986) found means of 26.7 (s.d. = 6.3) and 20.9 (s.d. = 5.3) in normal and depressed samples, respectively.

Reliability

Internal Consistency

Alpha coefficients range from .83 to .87 in the samples reported above.

Test–Retest

Test–retest reliabilities of .66 (males) and .68 (females) after 2 weeks were reported by Rorer and Goldberg (1965).

Validity

Convergent

The ESD is robust in that it correlates highly with scales from a variety of content areas if they too were assembled from items with extreme desirability ratings (Edwards, 1970; Jackson & Messick, 1962). For example, the SD correlates .71 with the Desirability scale of Jackson's (1967) PRF, which comprises items of extreme desirability chosen from diverse personality domains (Holden & Fekken, 1989).

Discriminant

Early critics alleged that, because 22 of the 39 items overlapped with Taylor's (1953) anxiety scale, the SD was simply another anxiety measure. Similarly, Crowne and Marlowe (1960) complained that the scale was intrinsically confounded with psychopathology because many of the items referred to psychological distress. Edwards and Walsh (1964) responded by showing that the pattern of correlations with other measures was unchanged when the psychopathology items were replaced.

Edwards distinguished the construct measured by the SD from tendencies to lie deliberately as measured by impression management scales (Edwards, 1957, 1970). This conceptual distinction has been clearly sustained by the data (Edwards & Walsh, 1964; Paulhus, 1984; Wiggins, 1964).

A major controversy was stirred by reports of a very high correlation between SD and the first factor of the MMPI (e.g., Jackson & Messick, 1962). To some observers this correlation suggested that the MMPI assesses SDR instead of psychopathology. Block (1965), however, argued forcefully that the SD reflects a substantive trait, namely, ego resiliency. Although most commentators agree that the SD reflects a more general disposition, the precise nature of the disposition remains moot (Paulhus, 1986).

Location

Because the items are taken directly from the MMPI (a copyrighted instrument), only a few sample items can be provided. However, the 39 MMPI booklet numbers are listed so that the full scale may be assembled by the reader. The MMPI is available from University of Minnesota Press, Minneapolis, Minnesota 55455.

Comments

Despite his many reports on the SD, Edwards has said very little about the nature of the underlying construct. Even in his most recent comment (Edwards *et al.*, 1988), he has

adhered closely to an operational definition: the tendency to respond desirably to the sort of item in his scale. Because the construct is not well defined, it is difficult to marshal evidence for it. As noted in the literature review, the SD falls clearly on the first factor of SDR, indicating an honest form of positivistic bias.

Edwards's original implication that a high correlation with SD invalidates a self-report measure is no longer tenable. In fact, controlling for SD may actually reduce the validity of adjustment-related measures (e.g., Kozma & Stones, 1988; McCrae, 1986).

Fortunately, Edwards's colleagues (L. K. Edwards & Clark, 1987) have finally published his alternative version (Edwards, 1963)—one comprising nonpsychopathology items. The scale appears to perform similarly to the original, thereby challenging certain critiques going back 30 years (Edwards, Edwards, & Clark, 1988). Nevertheless, the new version has yet to be subjected to scrutiny by other researchers.

Edwards Social Desirability Scale

Sample Items

True	False	1. I am happy most of the time. (T)
True	False	2. My hands and feet are usually warm enough. (T)
True	False	3. No one cares much what happens to you. (F)
True	False	4. I sometimes feel that I am about to go to pieces. (F)

Complete Scale

The MMPI booklet numbers for the 9 items keyed "True" are as follows: 7, 18, 54, 107, 163, 169, 257, 371, 528. The 30 items keyed "False" are 32, 40, 42, 43, 138, 148, 156, 158, 171, 186, 218, 241, 245, 247, 252, 263, 267, 269, 286, 301, 321, 335, 337, 352, 383, 424, 431, 439, 549, 555.

Marlowe–Crowne Social Desirability Scale (MCSD)

(Crowne and Marlowe, 1960)

Variable

Although Crowne and Marlowe (1960) originally constructed the MCSD to be a measure of SDR in self-reports, their subsequent research on the construct convinced them that the scale was tapping a more general motive: They dubbed it *need for approval* (Crowne & Marlowe, 1964).⁴ In the most recent statement, Crowne (1979) refined the concept to be an avoidance of disapproval.

Description

Crowne and Marlowe (1960) set out to build an SDR measure that improved upon the Edwards scale. Noting that Edwards's items were largely pathological in content, they

⁴Although Crowne was the senior author on both reports, the scale itself was labeled the Marlowe–Crowne scale, presumably to balance the credit.

focused instead on ordinary personal and interpersonal behaviors. Fifty such items were assembled and reduced to 33 by item analyses and ratings of experienced judges. The correlations with MMPI scales were still sizable, but not as high as those shown by the Edwards scale (e.g., Katkin, 1964).

The 33 items describe either (a) desirable but uncommon behaviors (e.g., admitting mistakes) or (b) undesirable but common behaviors (e.g., gossiping). Respondents are asked to respond "True" or "False" to 18 items keyed in the true direction and 15 in the false direction. Hence, scores range from 0 to 33, with higher scores representing higher need for approval.

Samples

Crowne and Marlowe (1964) reported a mean of 15.5 (s.d. = 4.4) in a sample of 300 college students. In a more recent study of 100 students, Paulhus (1984) reported means of 13.3 (s.d. = 4.3) and 15.5 (s.d. = 4.6) in anonymous and public disclosure conditions, respectively. In a sample of 503 students, Tanaka-Matsumi and Kameoka (1986) reported means of 14.0 and 12.3 for normal and depressed respondents, respectively. In a sample of 650 Peace Corps volunteers (90% college graduates), Fisher (1967) found means of 16.1 (s.d. = 6.8) and 16.4 (s.d. = 6.5) for males and females, respectively.

Reliability

Internal Consistency

Alpha coefficients ranged from .73 to .88 in the samples reported above.

Test–Retest

Crowne and Marlowe (1964) reported a test–retest correlation of .88 over 1 month. Fisher (1967) reported a value of .84 over a 1-week interval.

Validity

Convergent

The scale, as published in 1960, was intended as a measure of SDR in self-reports. A series of studies, summarized in Crowne and Marlowe (1964), uncovered a broad range of correlates suggesting the existence of an underlying motivational construct, namely, need for approval. For example, evidence showed that, compared to low scorers, high scorers on the MCSD respond more to social reinforcement, inhibit aggression, and are more amenable to social influence. Their task performance is more influenced by the evaluations of others. They prefer low-risk behaviors and avoid the evaluations of others, even when there is as much possibility for positive as for negative evaluation (for reviews of the research, see Crowne, 1979; Millham & Jacobson, 1978; Strickland, 1977).

Discriminant

As noted in the introduction, the MCSD falls primarily on the second SDR factor, showing only low to moderate correlations with such measures as Edwards SD and Self-Deceptive Enhancement.

Location

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Comments

The MCSD continues to sustain a dual existence as an SDR scale and a measure of the approval-dependent personality. Both interpretations are consistent with analyses showing the scale taps predominantly the second factor of SDR, that is, impression management (Paulhus, 1984).

One review (Strickland, 1977) was generally supportive of the need for approval construct, but recommended the label "approval motivation." Millham and Jacobson (1978) seem to prefer "evaluative dependence." The original prefix "need" was fashionable when the scale was developed but now seems presumptuous.

In addition, the weight of evidence has gradually shifted the interpretation to avoiding disapproval, rather than seeking approval, as implied by the original label (Allaman *et al.*, 1972; Crowne, 1979; Millham & Jacobson, 1978). Finally, some work has suggested that the attribution and denial items may be tapping distinct constructs (Millham, 1974; Paulhus & Reid, in press; see also Roth, Snyder, & Pace, 1986).

Use of the MCSD scale as a measure of situational demand is well-supported: Several studies have demonstrated its sensitivity to various audience effects (Davis & Cowles, 1989; Paulhus, 1984). Such effects, however, do not prove that subjects *consciously* modified their self-presentations.

More controversial is the question of whether high MCSD scores predict a proneness to dissimulation. A classic supporting example is Kiecolt-Glaser and Murray (1980): After an assertiveness training program, high MCSD scorers rated themselves as more assertive than low scorers although the program trainers rated them as *less* assertive. Other evidence suggests that high scorers will actually lie for reasons related to social approval (Jacobson, Berger, & Millham, 1970), but there is no clear evidence that they will lie for other reasons.

A complicating factor in interpreting certain studies is that, according to their spouses, high MCSD scorers actually do possess such desirable qualities as good adjustment, friendliness, and openness to experience (McCrae & Costa, 1983). Nonetheless, correlations in that study suggest that high MCSD scorers may further exaggerate their claims to such good qualities. A further complication is that high MCSD scorers also possess an honest demeanor: That is, judges tend to believe them and trust them even when they are instructed to lie (Riggio, Salinas, & Tucker, 1988). Indeed there is some evidence for a self-deceptive component (Millham & Kellogg, 1980; Weinberger, in press).

Marlowe-Crowne Scale

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you.

- T F 1. Before voting I thoroughly investigate the qualifications of all the candidates.

- T F 2. I never hesitate to go out of my way to help someone in trouble.
- T F *3. It is sometimes hard for me to go on with my work if I am not encouraged.
- T F 4. I have never intensely disliked anyone.
- T F *5. On occasion I have had doubts about my ability to succeed in life.
- T F *6. I sometimes feel resentful when I don't get my way.
- T F 7. I am always careful about my manner of dress.
- T F 8. My table manners at home are as good as when I eat out in a restaurant.
- T F *9. If I could get into a movie without paying and be sure I was not seen, I would probably do it.
- T F *10. On a few occasions, I have given up doing something because I thought too little of my ability.
- T F *11. I like to gossip at times.
- T F *12. There have been times when I felt like rebelling against people in authority even though I knew they were right.
- T F 13. No matter who I'm talking to, I'm always a good listener.
- T F *14. I can remember "playing sick" to get out of something.
- T F *15. There have been occasions when I took advantage of someone.
- T F 16. I'm always willing to admit it when I make a mistake.
- T F 17. I always try to practice what I preach.
- T F 18. I don't find it particularly difficult to get along with loud-mouthed, obnoxious people.
- T F *19. I sometimes try to get even, rather than forgive and forget.
- T F 20. When I don't know something I don't at all mind admitting it.
- T F 21. I am always courteous, even to people who are disagreeable.
- T F *22. At times I have really insisted on having things my own way.
- T F *23. There have been occasions when I felt like smashing things.
- T F 24. I would never think of letting someone else be punished for my wrongdoings.
- T F 25. I never resent being asked to return a favor.
- T F 26. I have never been irked when people expressed ideas very different from my own.
- T F 27. I never make a long trip without checking the safety of my car.
- T F *28. There have been times when I was quite jealous of the good fortune of others.
- T F 29. I have almost never felt the urge to tell someone off.
- T F *30. I am sometimes irritated by people who ask favors of me.

- | | | |
|---|---|----------------------------------------------------------------------------------------|
| T | F | 31. I have never felt that I was punished without cause. |
| T | F | *32. I sometimes think when people have a misfortune they only got what they deserved. |
| T | F | 33. I have never deliberately said something that hurt someone's feelings. |

Note: Items marked with an asterisk are keyed negatively.

MMPI Lie (L) Scale

(Meehl & Hathaway, 1946)

Variable

The L scale was designed to identify respondents who are deliberately trying to appear socially desirable while completing the MMPI.

Description

The scale comprises 15 statements about attitudes and practices that are socially undesirable but common. Topic areas include minor dishonesties, aggression, bad thoughts, and weaknesses of character. As on the MMPI as a whole, the response format is True-False. For all items, "False" is the response scored as a lie. In current usage, scores of 8 or above are considered suggestive of purposeful self-presentation (Greene, 1980).

Samples

Hathaway and McKinley (1951) reported means of 4.2 (s.d. = 2.6) and 4.5 (s.d. = 2.6) for males and females, respectively. In a more recent sample of 765 college students, Goldberg (1972) reported means of 2.5 (s.d. = 1.9) and 2.7 (s.d. = 1.8) for males and females, respectively. In a massive sample of 50,000 medical outpatients, Swenson, Pearson, and Osbourne (1973) reported means of 4.2 (s.d. = 2.3) and 4.8 (s.d. = 2.3) for males and females, respectively.

Reliability

Internal Consistency

Gocka (1965) reported an alpha coefficient of .72 on a patient sample. Paulhus (1984) reported an alpha value of .60 on a student sample.

Test-Retest

Test-retest correlations for intervals of up to 1 week range from .70 to .85. For intervals of 1 year or more, correlations range from .35 to .60 (Greene, 1980; Rorer & Goldberg, 1965).

Validity

Convergent

Evidence for concurrent validity is available from studies showing high correlations with similar constructs, for example, Eysenck's Lie scale (Paulhus, 1986) and the Marlowe-Crowne scale (e.g., Edwards & Walsh, 1964). As noted in the introduction, the L scale loads on both factors of SDR but primarily on the second factor, impression management.

No claim has been made that high scores should predict lying outside of self-reports. In the only known laboratory study, high scorers performed better than lows in a stressful situation (Burish & Houston, 1976).

Discriminant

The measure shows low correlations with measures loading on the first factor of SDR. For example, correlations with the K scale, the alternative SDR measure from the MMPI, range from low to moderate (Dahlstrom, Welsh, & Dahlstrom, 1972).

Location

Hathaway, S. R., & McKinley, J. C. (1951). *MMPI manual*. New York: Psychological Corporation.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook* (Vol. 1). Minneapolis: University of Minnesota Press.

Comments

It was soon apparent to the constructors of the L scale that the desirability demand of the items was less than subtle: "The L score was a trap for the naive subject but easily avoided by more sophisticated subjects" (Meehl & Hathaway, 1946). It is obvious to a sophisticated test-taker that, even if one is trying to appear desirable, it is unrealistic to deny such ubiquitous attributes.

Given its negative correlation with intelligence, some have construed the L scale as a measure of psychological sophistication. College-educated persons and those of higher

MMPI Lie Scale

Sample Items

- At times I feel like swearing. (F)
TRUE FALSE
- I get angry sometimes. (F)
- Sometimes when I am not feeling well I am cross. (F)

Complete Scale

The MMPI booklet numbers for the 15-item Lie scale are 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 165, 195, 225, 255, 285. All items are keyed negatively, that is, one point is assigned for each "False."

socioeconomic status rarely score above 4 (Dahlstrom *et al.*, 1972). When they do, it may suggest a gross lack of insight into their own behavior (Greene, 1980).

MMPI K Scale

(Meehl & Hathaway, 1946)

Variable

The K scale was designed as a subtle measure of SDR on the MMPI. It is used to identify persons with psychopathology whose MMPI protocols appear normal.

Description

The 30 items were selected empirically. First, MMPI responses from normals were compared to those of persons with known psychopathology who scored as normals on the clinical scales. This procedure yielded 22 discriminating items. Depressed and schizophrenic patients, however, scored low on these items. Therefore, eight items were added to differentiate these two groups from normals. In current usage, scores of 16 or above are said to suggest invalid MMPI protocols (Greene, 1980).

Samples

In the original MMPI sample of 610 normals (a cross section of Minnesota residents), McKinley *et al.* (1948) reported means of 12.8 (s.d. = 5.6) and 12.1 (s.d. = 5.1) for males and females, respectively. In their mixed sample of 968 psychiatric cases, the means were 14.6 (s.d. = 5.9) and 14.3 (s.d. = 5.2) for males and females, respectively. In their sample of 100 university students, the means were 16.1 (s.d. = 5.1) and 15.7 (s.d. = 5.0) for males and females, respectively.

In a more recent sample of college students, Goldberg (1972) reported means of 15.4 (s.d. = 4.7) and 15.5 (s.d. = 4.3) for males and females, respectively. In a massive sample of 50,000 medical outpatients, Swenson *et al.* (1973) reported means of 15.4 (s.d. = 4.9) and 15.5 (s.d. = 4.8) for males and females, respectively.

Reliability

Internal Consistency

Gocka (1965) reported an alpha coefficient of .82 on a patient sample.

Test-Retest

Correlations range from .78 to .92 for an interval up to 2 weeks and range from .52 to .67 for intervals from 8 months to 3 years (Greene, 1980; Rorer & Goldberg, 1965).

Validity

Convergent

According to the original test constructors, the scale "was not assumed to be measuring anything which in itself is of psychiatric interest" (Meehl & Hathaway, 1946, p. 544). The

validation of the scale was considered to rest on its value as a correction factor. That is, controlling other measures for K should improve their predictive validity.

In this respect, the validation evidence is weak. A few studies have shown improved validity of the MMPI scales after correcting scores as recommended in the MMPI manual (e.g., Wooten, 1984). Most studies have shown, if anything, decreases in validity (e.g., Heilbrun, 1963; McCrae *et al.*, 1989; Yonge, 1966).

The validity of K as a measure of defensiveness is better supported but appears to vary according to the type of respondent. In maladjusted college students, there is evidence that K indexes defensiveness (e.g., Heilbrun, 1961; Reis, 1966). In normal college students, K appears to tap a healthy positive self-image (e.g., McCrae *et al.*, 1989; Yonge, 1966).

Location

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*, 525-564.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook* (Vol. 1.) Minneapolis: University of Minnesota Press.

Comments

As noted in the introduction, the K scale falls on the first factor of SDR. This association is consistent with the original conception of the K scale as a subtle measure of desirable responding. At the same time, this association suggests that at least some high scores result from a positive bias in self-image.

MMPI K Scale

Sample Items

1. I like to let people know where I stand on things. (T)
TRUE FALSE
2. I have very few quarrels with members of my family. (T)
3. People often disappoint me. (F)

Complete Scale

The MMPI booklet numbers for the items keyed "False" are 30, 39, 71, 89, 124, 129, 134, 138, 142, 148, 160, 170, 171, 180, 183, 217, 234, 267, 272, 296, 316, 322, 374, 383, 397, 398, 406, 461, 502. The sole item keyed "True" is MMPI booklet number 96.

CPI Good Impression (Gi) Scale*(Gough, 1952)***Variable**

The Gi scale was designed to measure what people say about themselves when trying to create an extremely favorable impression (Gough, 1987, p. 36).

Description

Following Ruch (1942), test development involved contrast groups. Subjects first took an experimental booklet of items under normal circumstances and then repeated the testing with "good impression" instructions: "Try to give just as favorable an impression of yourself as you would if you were actually applying for an important position, or were trying to create a very favorable impression. . ."

The items tested included some adopted from Ruch (1942) and others newly written to measure impression management. The 40 best-differentiating items were included on Gough's (1957) CPI and five were modified for the revised CPI (Gough, 1987). Scores can range from 0 to 40 with scores above 30 suggestive of faking good.

Samples

Gough (1987) reported means and standard deviations for a wide variety of samples including 4126 college students (18.5, s.d. = 5.9), 100 nurses (18.6, s.d. = 5.5), and 345 prison inmates (17.9, s.d. = 7.0).

Reliability*Internal Consistency*

Gough (1987) reported alpha coefficients of .77 for both male and female college students.

Test-Retest

Gough (1987) reported test-retest correlations of .68 after 1 year for both male and female high school students.

Validity*Convergent*

A total of 400 CPI respondents were rated by their spouses using Q-sorts. The four Q-sort items showing the largest positive correlations with respondents' Gi scores were: (1) A conscientious and serious-minded person, (2) Well-organized, capable, patient, and industrious; values achievement, (3) Gentle, considerate, and tactful in dealing with others, and (4) Gets along well with others; able to "fit in" easily in most situations. In addition, 793 respondents were rated on Q-sorts by trained assessors. The four highest correlating

items were (1) Is fastidious, (2) Favors conservative values in a variety of areas, (3) Is a genuinely dependable and responsible person, (4) Tends toward overcontrol of needs and impulses; binds tensions excessively; delays gratification unnecessarily. These and other validity data suggest that the high scorer is a highly controlled individual who behaves in a socially conventional manner.

Location

Gough, H. G. (1987). *California Psychological Inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.

Comments

The substantial validity data on the Gi highlight the dualistic nature of SDR measures. According to people who should know (spouses, peers, interviewers), the high scorer actually is a socially desirable person in being dependable, industrious, and cooperative.

However, the raters also see the high scorer as overcontrolled, suggesting an unwillingness to acknowledge undesirable qualities. This defensiveness is borne out by the fact that to score high on the Gi the respondent had to claim desirable qualities well beyond those validated by the judges (e.g., cultured interests, social skills). Other evidence of the high scorer's tendency to put the best foot forward is that interviewers rated him/her as well groomed, well dressed, and polite.

This dualism in the target construct is handled by giving a substantive interpretation to scores up to the cutoff point of 30, after which respondents are assumed to be faking good (Lanning, 1989).

The qualities measured by a scale developed through role-playing instructions depend wholly on the instructions given to the experimental group. As the reader may see above, the instructions used to select Gi items mentioned the job interview, thereby targeting the conventional, dependable, industrious, and cooperative types. This persona is, of course, only one of many possible good impressions.

CPI Gi Scale**Sample Items**

1. I always follow the rule: business before pleasure. (T)
TRUE FALSE
2. I have never deliberately told a lie. (F)
3. I enjoy hearing lectures on world affairs. (T)

Complete Scale

The CPI booklet numbers for the items keyed "True" are 14, 103, 127, 133, 140, 165, 195, 222, 254. The items keyed "False" are 10, 30, 34, 38, 42, 44, 48, 56, 66, 70, 78, 81, 91, 101, 102, 109, 120, 150, 153, 159, 170, 178, 203, 207, 231, 238, 248, 262, 268, 273, 289, 293.

Balanced Inventory of Desirable Responding (BIDR)

(Paulhus, 1984, 1988)

Variable

The BIDR measures two constructs: self-deceptive positivity (the tendency to give self-reports that are honest but positively biased) and impression management (deliberate self-presentation to an audience).

Description

The BIDR is a descendant of the Self- and Other-Deception Questionnaires developed by Sackeim and Gur (1978). The original self-deception items were rationally developed on the assumption that individuals with a propensity for self-deception tend to deny having psychologically threatening thoughts or feelings. The threats were based on psychoanalytic theory (e.g., hating one's parents, enjoying one's bowel movements, having sexual fantasies). In contrast, the more recent version of the scale (Paulhus, 1988) emphasizes exaggerated claims of positive cognitive attributes (overconfidence in one's judgments and rationality). Thus the focus has shifted from ego defense to ego enhancement. Given that the newer measure of self-deception is presented here, the psychometric information reported below applies only to that version.

The impression management items were rationally developed on the assumption that some respondents systematically overreport their performance of a wide variety of desirable behaviors and underreport undesirable behaviors. Because the claims involve overt behaviors (e.g., I always pick up my litter), any distortion is presumably a conscious lie.

The 40 BIDR items are stated as propositions. Respondents rate their agreement with each statement on a seven-point scale. The scoring key is balanced. After reversing the negatively keyed items, one point is added for each extreme response (6 or 7). Hence, total scores on SDE and IM can range from 0 to 20. This scoring ensures that high scores are attained only by subjects who give exaggeratedly desirable responses. All 40 items may be summed to yield an overall measure of SDR that correlates highly with the MCSD. (An extended version including 20 denial items is also available.)

Samples

Self-Deception

In a large sample of 884 religious adults, Quinn (1989) found means of 7.6 (s.d. = 3.1) and 7.3 (s.d. = 3.1) for males and females, respectively. In a sample of 433 college students, Paulhus (1988) reported corresponding means of 7.5 (s.d. = 3.2) and 6.8 (s.d. = 3.1).

Impression Management

Quinn (1989) reported male and female means of 7.3 (s.d. = 3.1) and 8.9 (s.d. = 3.2) in a sample of 884 religious adults. In a sample of 433 college students, Paulhus (1988) reported means of 4.3 (s.d. = 3.1) and 4.9 (s.d. = 3.2) for males and females, respectively. In a sample of 100 college students, Paulhus (1984) reported an overall mean of

11.9 (s.d. = 4.5) in a public disclosure condition. In a sample of 48 members of alcoholics anonymous, Mellor, Conroy, and Masteller (1986) reported a mean of 11.2 (s.d. = 4.9).

Reliability

Internal Consistency

In the studies reported above, values of coefficient alpha range from .68 to .80 for the SDE and from .75 to .86 for the IM scale. When all 40 items are summed as a measure of SDR, the alpha is .83 (Paulhus, 1988).

Test-Retest

Paulhus (1988) reported test-retest correlations over a 5-week period of .69 and .65 for the SDE and IM scale, respectively.

Validity

The sum of all 40 BIDR items shows concurrent validity as a measure of SDR in correlating .71 with the Marlowe-Crowne scale (Paulhus, 1988) and .80 with the Multidimensional Social Desirability Inventory of Jacobson, Kellogg, Cauce, and Slavin (1977).

Convergent: Self-Deception

In general, measures of self-deception show concurrent validity in correlating strongly with other first factor SDR measures (see introduction). Paulhus (1988) found that the SDE measure provided here correlates positively with the following traditional measures of defense and coping: (1) repressive style as measured by Byrne's R-S scale ($r = .51$), (2) reversal, as measured by Ihlavich and Gleser's (1986) Defense Mechanisms Inventory ($r = .34$), and (3) positive re-appraisal ($r = .44$), distancing ($r = .33$), and self-controlling ($r = .39$) as measured with the Ways of Coping scale (Folkman, Lazarus, Dunkel-Schetter, DeLongis, & Gruen, 1986).

Several experimental studies have supported the construct validity of the SDE. After a failure experience, high self-deception subjects were more likely than lows to show a self-serving bias (Paulhus, 1988). High self-deception subjects also showed more illusion of control, belief that they are safe drivers, and proneness to love (Paulhus & Reid, in press) and to intrinsic religiosity (Leak & Fish, 1989). High scorers also show excessive confidence in memory judgments and more hindsight bias; they also claim familiarity with non-existent products (Paulhus, 1988).

All these mechanisms may contribute to the positive adjustment reported by high SDE subjects including high self-esteem as well as low neuroticism, depression, empathic distress, and social anxiety (Paulhus & Reid, in press). Note that all these measures of adjustment have been validated in the past by clinical judgment, behavioral measures, and/or peer-ratings.

Convergent: IM Scale

As noted in the introduction, the IM scale correlates highly with a cluster of measures traditionally known as lie scales (e.g., Eysenck's Lie scale, MMPI Lie scale) and role-playing measures (e.g., Wiggins' Sd, Gough's Gi). Correlations with the MCSD and

agreeableness and conscientiousness ratings (Paulhus, 1988) suggest that a social approval motive underlies anonymous responses.

The IM scale is particularly responsive to demands for impression management. For example, in a comparison of six SDR measures, the IM scale showed the largest increase from private to public conditions (Paulhus, 1984). Lautenschlager and Flaherty (in press) showed that IM, but not SDE, was sensitive to test administration conditions (paper and pencil vs. computer; public vs. private).

Discriminant

Measures of self-deception and impression management show discriminant validity in forming separate factors in factor analyses (Paulhus, 1984, 1988). Earlier versions of the self-deception measure showed positive correlations with impression management ranging from .35 to .65, depending on the situational demand for self-presentation. The version presented here, however, exhibits much lower correlations, ranging from .05 to .40. Note that males score higher than females on self-deception, but lower on impression management.

Location

Paulhus, D. L. (1988). *Assessing self deception and impression management in self-reports: the Balanced Inventory of Desirable Responding*. (Manual available from the author at the Department of Psychology, University of British Columbia, Vancouver, B.C., Canada V6T 1Y7.)

Comments

The predecessors of these measures, the Self-Deception and Other-Deception Questionnaires, were first described in Sackeim and Gur (1978), although the items have never been published. Subsequently, Paulhus (1984) refined the measures and integrated them into one inventory. The two major refinements were (a) writing reversals to balance the keys, and (b) replacing the psychopathology items. The latter refinement eliminated any spurious correlation with psychopathology measures. Five preliminary versions of the BIDR preceded the version presented here (Paulhus, 1988). A French language version is also available (Sabourin, Bourgeois, Gendreau, & Morval, 1989).

A major feature of the BIDR is the provision for separate measures of the two major SDR factors, self-deceptive enhancement and impression management. It is often critical to know which component is responsible for a correlation observed between SDR and some other variable. In addition, the dichotomous scoring procedure (assigning points only for *extremely* desirable responses) provides some assurance that style rather than content is being tapped. IM is more likely to tap style as anonymity decreases.

Note that substantial correlations are observed between SDE and measures of adjustment even though the content of the SDE measure is free of psychopathology. These findings suggest that self-deceptive positivity is intrinsically linked to the adjusted personality, consistent with current views of adjustment (Alloy & Abramson, 1979; Taylor & Brown, 1988). Research is required on the personality of extreme scorers: Peer-raters may not see them as well-adjusted as they see themselves. They may also snap under stress.

Validation of a measure of self-deception is constrained by the uncertain status of the construct (see Lockard & Paulhus, 1988). Conceptually similar labels for the construct are available, for example lack of insight, overconfidence, or dogmatism. Whatever the label,

it is clear that the SDE scale is tapping a specific form of SDR, one that is less subject to purposeful manipulation than measures in the impression management category.

BIDR Version 6—Form 40

Using the scale below as a guide, write a number beside each statement to indicate how much you agree with it.

1 2 3 4 5 6 7
 NOT TRUE SOMEWHAT TRUE VERY TRUE

- ___ 1. My first impressions of people usually turn out to be right.
- ___ *2. It would be hard for me to break any of my bad habits.
- ___ 3. I don't care to know what other people really think of me.
- ___ *4. I have not always been honest with myself.
- ___ 5. I always know why I like things.
- ___ *6. When my emotions are aroused, it biases my thinking.
- ___ 7. Once I've made up my mind, other people can seldom change my opinion.
- ___ *8. I am not a safe driver when I exceed the speed limit.
- ___ 9. I am fully in control of my own fate.
- ___ *10. It's hard for me to shut off a disturbing thought.
- ___ 11. I never regret my decisions.
- ___ *12. I sometimes lose out on things because I can't make up my mind soon enough.
- ___ 13. The reason I vote is because my vote can make a difference.
- ___ *14. My parents were not always fair when they punished me.
- ___ 15. I am a completely rational person.
- ___ *16. I rarely appreciate criticism.
- ___ 17. I am very confident of my judgments.
- ___ *18. I have sometimes doubted my ability as a lover.
- ___ 19. It's all right with me if some people happen to dislike me.
- ___ *20. I don't always know the reasons why I do the things I do.
- ___ *21. I sometimes tell lies if I have to.
- ___ 22. I never cover up my mistakes.
- ___ *23. There have been occasions when I have taken advantage of someone.
- ___ 24. I never swear.
- ___ *25. I sometimes try to get even rather than forgive and forget.
- ___ 26. I always obey laws, even if I'm unlikely to get caught.
- ___ *27. I have said something bad about a friend behind his or her back.
- ___ 28. When I hear people talking privately, I avoid listening.

- ___ *29. I have received too much change from a salesperson without telling him or her.
- ___ 30. I always declare everything at customs.
- ___ *31. When I was young I sometimes stole things.
- ___ 32. I have never dropped litter on the street.
- ___ *33. I sometimes drive faster than the speed limit.
- ___ 34. I never read sexy books or magazines.
- ___ *35. I have done things that I don't tell other people about.
- ___ 36. I never take things that don't belong to me.
- ___ *37. I have taken sick-leave from work or school even though I wasn't really sick.
- ___ 38. I have never damaged a library book or store merchandise without reporting it.
- ___ *39. I have some pretty awful habits.
- ___ 40. I don't gossip about other people's business.

Items 1–20 assess SDE; items 21–40 assess IM. Add one point for every "6" or "7" (minimum = 0; maximum = 20).

*, Items keyed in the "False" (negative) direction.

Responding Desirably on Attitudes and Opinions (RD-16)

(Schuessler, Hittle, & Cardascia, 1978)

Variable

This measure of SDR was specially designed to detect socially desirable responding in attitude and opinion surveys of the general population.

Description

The scale development involved several stages. A set of 270 items was taken from over 100 tests of attitudes, morale, and related concepts. These were rated for desirability and items showing neutral ratings, above average variance, or interactions between race and education were discarded. Sixteen items were selected rationally to cover the widest range of topics. The scale was normed in a national probability sample of 1522 adults.

The 16-item scale comprises eight pairs, one pair from tests of dejection, social estrangement, social opportunism, trust, social contentment, anomie, expediency, and self-determination. Each pair (and therefore, the entire set) is key-balanced. The subject is asked to agree or disagree with each item. Possible scores range from 0 to 16 with higher scores indicating more desirable responding.

Samples

In their national probability sample of 1522 adults, Schuessler *et al.* (1978) reported an overall mean of 12.7 (s.d. = 2.4). On a shortened 10-item version, Krebs and Schuessler

(1987) reported means of 7.9 (s.d. = 1.6) and 7.2 (s.d. = 1.6) on American and German samples, respectively.

Reliability

Internal Consistency

Schuessler *et al.* (1978) reported an overall alpha coefficient of .64 in the national probability sample.

Test-Retest

No test-retest reliabilities are available.

Validity

Convergent

Concurrent validity is supported by a correlation of .55 with a 10-item version of Jackson's (1967) Desirability scale.

Discriminant

According to Schuessler *et al.* (1978), discriminant validity is supported by low intercorrelations with the Marlowe-Crowne scale ($r = .07$ to $.16$). Low to moderate ($-.18$ to $-.36$) negative correlations were found with measures of acquiescence (Krebs & Schuessler, 1987; Schuessler *et al.*, 1978).

Location

Schuessler, K., Hittle, D., & Cardascia, J. (1978). Measuring responding desirably with attitude-opinion items. *Social Psychology*, **41**, 224–235.

Comments

Several attributes make the RD-16 appropriate for use in attitude and opinion surveys. First, the items were drawn from a wide pool of general attitude and opinion measures. Second, the items were selected from ratings done by a cross section of adults rather than the usual college sophomores. Third, the items were screened so that none would differ in desirability across race and education.

It is notable that, although the items were taken from attitude and opinion surveys, the statistical selection procedures yielded personality-oriented items like those on other SDR measures. The measure falls clearly on the first factor of SDR (see introductory material at the beginning of this chapter). A major feature is the national probability sample of norms broken down by social status categories. A German language version is also available (Krebs & Schuessler, 1987). Note that the items below should be randomized for presentation: They are listed below in all-desirable then all-undesirable order.

RD-16

Please circle "A" or "D" to indicate whether you agree or disagree with the following statements. Do not omit any items.

- A D 1. I find that I can help others in many ways.
 A D 2. I feel that I am better off than my parents were at my age.
 A D 3. In spite of many changes, there are still definite rules to live by.
 A D 4. One can always find friends if he [one] tries.
 A D 5. Anyone can raise his standard of living if he [one] is willing to work at it.
 A D 6. Most people really believe that honesty is the best policy.
 A D 7. In general, I am satisfied with my lot in life.
 A D 8. People will be honest with you as long as you are honest with them.
 A D *9. It is difficult to think clearly about right and wrong these days.
 A D *10. Many people are friendly only because they want something from you.
 A D *11. If the odds are against you, it's impossible to come out on top.
 A D *12. At times I feel that I am a stranger to myself.
 A D *13. The future looks very bleak.
 A D *14. I often feel that no one needs me.
 A D *15. I am so fed up that I can't take it any more.
 A D *16. To get along with people one must put on an act.

Note: *, Items keyed in "False" (negative) direction. Nonsexist wording is suggested in brackets.

Children's Social Desirability Scale (CSD)

(Crandall, Crandall, & Katkovsky, 1965)

Variable

Modeled after the Marlowe-Crowne scale, the CSD assesses SDR in children as motivated by a need for approval. Subsequently, the construct was reinterpreted as fear of disapproval (Crandall, 1966).

Description

The scale contains 48 statements in True-False format with 26 items keyed true. Much of the item content follows that of the Marlowe-Crowne items. Additional items involve

child-specific content (e.g., "Sometimes I want to do things my parents think I am too young to do.") or items that are worded in children's language (e.g., "Sometimes I wish I could just mess around instead of having to go to school."). Possible scores range from 0 to 48 with high scores indicating a fear of disapproval.

Samples

In a total sample of 956 grade school and high school students, Crandall *et al.* (1965) reported means monotonically decreasing from 29.3 (s.d. = 10.4) in Grade 3 to 12.7 (s.d. = 7.6) in Grade 12.

Reliability*Internal Consistency*

Crandall *et al.* (1965) reported corrected split-half reliabilities ranging from .82 to .95.

Test-Retest

Crandall *et al.* (1965) reported a test-retest correlation of .85 after a 1-month interval. Allaman *et al.* (1972) reported test-retest correlations of .90 after 1 month and of .43 and .19 over 3 years for males and females, respectively.

Validity*Convergent*

Correlations of .78 and .51 were found between scores on the CSD and scores on the Marlowe-Crowne and Good Impression scales, respectively. High scorers were more religious (Crandall & Gozali, 1969) but had lower self-esteem (Crandall, 1966). In observational studies, high CSD scorers were less aggressive, participated less, avoided achievement activities (Crandall, 1966), and ate less candy in the presence of others (Staub & Sherk, 1970). Some evidence is less consistent with the construct (for a review, see Strickland, 1977).

Discriminant

No relevant information has been reported.

Location

Crandall, V., Crandall, V. J., & Katkovsky, W. A. (1965). A children's social desirability scale. *Journal of Consulting Psychology*, *29*, 27-36.
 The actual items are presented in Strickland (1977).

Comments

Crandall *et al.* (1965) also developed a version for younger children (Grades 3-5). The items are posed in question format and are usually presented via a recording for standardization purposes. In a sample of 43 fifth graders, this measure correlated .85 with the written true-false version (Brannigan, 1974).

The developmental data are provocative in suggesting that SDR tendencies result from maternal tendencies such as hostility, criticism, restrictiveness, punitiveness, and lack of encouragement. In addition, noncompliance and dominance in infancy were related to adult SDR (Allaman *et al.*, 1972).

Developmental stability is a critical issue. Test-retest correlations are high after 1 month but low after 3 years. Crandall *et al.* (1965) interpret the latter as developmental instability. Interestingly, the correlation with the Marlowe-Crowne noted earlier (.78) was after an average delay of several years. Although the sample size was very small ($n = 12$), Allaman *et al.* (1972) suggest that SDR may stabilize by adolescence.

CSD

This questionnaire lists a number of experiences that most children have at one time or another. Read each of these carefully. After you have read one, decide whether it does or does not fit you. If it does, put a T (for true) in front of the statement; if it doesn't, put an F (for false) in front of the statement.

- ___ 1. I always enjoy myself at a party.
- ___ *2. I tell a little lie sometimes.
- ___ 3. I never get angry if I have to stop in the middle of something I'm doing to eat dinner, or go to school.
- ___ *4. Sometimes I don't like to share my things with my friends.
- ___ 5. I am always respectful of older people.
- ___ 6. I would never hit a boy or girl who was smaller than me.
- ___ *7. Sometimes I do not feel like doing what my teachers want me to do.
- ___ 8. I never act "fresh" or "talk back" to my mother or father.
- ___ 9. When I make a mistake, I always admit I am wrong.
- ___ *10. I feel my parents do not always show good judgment.
- ___ 11. I have never felt like saying unkind things to a person.
- ___ 12. I always finish all of my homework on time.
- ___ *13. Sometimes I have felt like throwing or breaking things.
- ___ 14. I never let someone else get blamed for what I did wrong.
- ___ *15. Sometimes I say something just to impress my friends.
- ___ 16. I am always careful about keeping my clothing neat, and my room picked up.
- ___ 17. I never shout when I feel angry.
- ___ *18. Sometimes I feel like staying home from school even if I am not sick.
- ___ *19. Sometimes I wish that my parents didn't check up on me so closely.
- ___ 20. I always help people who need help.
- ___ *21. Sometimes I argue with my mother to do something she doesn't want me to do.
- ___ 22. I never say anything that would make a person feel bad.
- ___ 23. My teachers always know more about everything than I do.
- ___ 24. I am always polite, even to people who are not very nice.

- ___ *25. Sometimes I do things I've been told not to do.
- ___ 26. I never get angry.
- ___ *27. I sometimes want to own things just because my friends have them.
- ___ 28. I always listen to my parents.
- ___ 29. I never forget to say "please" and "thank you."
- ___ *30. Sometimes I wish I could just "mess around" instead of having to go to school.
- ___ 31. I always wash my hands before every meal.
- ___ *32. Sometimes I dislike helping my parents even though I know they need my help around the house.
- ___ 33. I never find it hard to make friends.
- ___ 34. I have never been tempted to break a rule or a law.
- ___ *35. Sometimes I try to get even when someone does something to me I don't like.
- ___ *36. I sometimes feel angry when I don't get my way.
- ___ 37. I always help an injured animal.
- ___ *38. Sometimes I want to do things my parents think I am too young to do.
- ___ *39. I sometimes feel like making fun of other people.
- ___ 40. I have never borrowed anything without asking permission first.
- ___ *41. Sometimes I get annoyed when someone disturbs something I've been working on.
- ___ 42. I am always glad to cooperate with others.
- ___ 43. I never get annoyed when my best friend wants to do something I don't want to do.
- ___ *44. Sometimes I wish that the other kids would pay more attention to what I say.
- ___ 45. I always do the right things.
- ___ *46. Sometimes I don't like to obey my parents.
- ___ *47. Sometimes I don't like it when another person asks me to do things for him.
- ___ *48. Sometimes I get mad when people don't do what I want.

Note: *, Items keyed in "False" (negative) direction.

Acquiescence

Acquiescence is the tendency to agree rather than disagree with propositions in general (e.g., Lentz, 1938). A few studies have examined the effects of test situation and item format (e.g., Schuman & Presser, 1981; Trott & Jackson, 1967), but the bulk of the research has addressed acquiescence as a response style. Some individuals, called

yeasayers, tend to agree with statements or say "yes" to questions; other individuals, called *naysayers*, tend to disagree with statements or say "no" to questions. Rather than a mechanical response to any question, this tendency is assumed to emerge when the subject is uncertain (Peabody, 1966). Acquiescence has often been viewed as an individual difference variable in its own right, a personality trait with conceptual links to conformity and impulsiveness (e.g., Couch & Keniston, 1960; Gough & Heilbrun, 1980; Messick, 1967).

A problem arises when a self-report instrument measures acquiescence as well as the construct it was designed to measure. For example, on many anxiety scales subjects are asked to indicate with a "yes" or "no" which anxiety-related symptoms they have experienced. The respondent who says "yes" to all the symptoms may indeed be a very anxious person. Alternatively, the respondent may merely be a yeasayer.

Some researchers have claimed that acquiescence can be a serious confound in self-reports of attitudes (Carr, 1971; Ray, 1983; Schuman & Presser, 1981), ability and achievement (Cronbach, 1946), personality (Jackson & Helmes, 1979), and psychopathology (Jackson & Messick, 1958). Moreover, acquiescence has been found to interact with social status variables such as race and education (Bachman & O'Malley, 1984; DeLamater & McKinney, 1982). In contrast, other researchers have concluded that acquiescence effects are insignificant (Gove & Geerken, 1977; Rorer, 1965; Rorer & Goldberg, 1965; Wright, 1975).

With a view to reconciling these conflicting data, Bentler, Jackson, and Messick (1971) distinguished two types of acquiescence: agreement acquiescence and acceptance acquiescence. *Agreement* acquiescence is the tendency to agree with (or give positive ratings) to all types of items, even an item and its own negation ("happy" and "not happy"). *Acceptance* acquiescence is the tendency to endorse all qualities (even apparently contrary ones) as true of one's self. This form is indicated by agreeing that one is both "happy" and "sad" and disagreeing that one is "not happy" and "not sad." Bentler *et al.* (1971) concluded that the effects of agreement acquiescence are insignificant, whereas acceptance acquiescence remains a problem. Block (1971) was skeptical.

There is little disagreement that acquiescence is more problematic in attitude and survey research than in personality assessment (Bentler *et al.*, 1971; Ray, 1983; Schuman & Presser, 1981). In survey research, the percentage agreement with an item (e.g., I favor capital punishment) is usually more critical than in personality items (e.g., I am friendly), in which relative agreement across items is the issue. Moreover, in many personality inventories, the items are simply trait adjectives, thereby simplifying the control of acquiescence. In sharp contrast, Schuman and Presser (1981) have shown that the complex statements required in much survey research are highly susceptible to acquiescence in agree-disagree, interrogative, or true-false format. Moreover, Ray's (1983) work suggests that acquiescence may be a generalized style across attitude scales.

Control

Despite continuing disagreement about the pervasiveness of acquiescence, most scale constructors now make an effort to balance the scoring key. Usually, half the items are keyed positively (a high rating indicates possession of the construct being assessed) and half the items are keyed negatively (a low rating indicates possession of the construct). In dichotomous formats (e.g., True-False), this procedure is equivalent to keying half the items true and half false.

This simple precaution controls the classical form of acquiescence (agreement acquiescence) because, to get an overall high score, the respondent must agree with many

items and disagree with many others. In other words, one cannot get a high score simply by yea-saying or nay-saying (for a cautionary note, see Wiggins, 1968).

It is more difficult to correct an imbalanced scale post hoc. If the correlations are high between positively and negatively keyed subtotals and their correlations with other variables are comparable, then one may safely combine the two. One could then differentially weight the positively and negatively keyed subtotals to simulate a balanced key (Winkler, Kanouse, & Ware, 1982). Partial correlation techniques have also been applied to remove acquiescence statistically (Webster, 1958).

According to Bentler *et al.* (1971), however, it is acceptance acquiescence that specifically requires controlling. Simply adding a negation (not happy) for each item worded as an assertion (happy) will not suffice.⁵ One must add conceptual opposites that are also worded as assertions. When collecting ratings on the personality trait "dominance," for example, one must also include the conceptually opposite trait, "submissive." In applications where the conceptual opposite is not clear (as in many survey items), preliminary studies may be necessary to find an appropriate assertion to match the original (for examples, see Schuman & Presser, 1981). Combining these matched options in a forced-choice format is even better. Only one known personality instrument has been designed to control both forms of acquiescence: the Multidimensional Social Desirability Inventory (Jacobson, Brown, & Ariza, 1983; Jacobson *et al.*, 1977).⁶ Note, however, that the instrument targets only one domain, socially desirable responding.

Measurement

A small number of instruments have been designed to measure individual differences in the tendency to acquiesce in self-reports. The Couch-Keniston (1960) agreement response scale, for example, was included in the first edition of the present volume. On the basis of empirical work (e.g., Rorer, 1965), however, such measures have fallen into disrepute and none is widely used. Moreover, the original proponents of the importance of acquiescence, Messick and Jackson, shifted their focus to acceptance acquiescence, a form that seems to be domain-specific. In short, none of the instruments claiming to measure general acquiescence tendencies can be recommended to the researcher. Therefore, no such measures are presented in this chapter.

A number of larger assessment batteries permit computation of an acquiescence index across all the items in the battery. Gough and Heilbrun's (1965) Adjective Check List, for example, permits calculation of the "checking factor," that is, the total number of adjectives checked as true of the self. This score is often factored out of subsequent analyses (e.g., Wiggins, 1979). Note that this procedure may eliminate some content unless one has administered the ACL in true-false format to ensure some response to each item. The MMPI permits detection of "all true" or "all false" protocols through computation of the Carelessness scale (Greene, 1980).

Several statistical techniques have been designed to separate the contributions of item content from both forms of acquiescence (Bramble & Wiley, 1974; Morf & Jackson, 1972). If, however, acceptance acquiescence is domain-specific, one may require a separate measure for each personality dimension or narrow battery. Indeed, one may even require separate measures for the positively and negatively keyed items within a scale (Paulhus & Reid, in press).

⁵Another problem with negating items may be a loss in validity (Holden *et al.*, 1985).

⁶Sample items of each type are: "I always keep my promises," "I daydream about sexual acts," "I do not always vote," and "I never feel worthless."

Extremity Response Bias (ERB)

Extremity response bias (ERB) is the tendency to use the extreme choices on a rating scale (e.g., 1s and 7s on a seven-point scale). Situational factors such as ambiguity (Shulman, 1973), emotional arousal, and speededness (Paulhus, 1987) induce temporary increases in ERB. The individual exhibiting a consistent ERB across time and stimuli may be said to have an extremity response style; low scorers on this construct may be said to have a moderacy response bias, tending to use the midpoint as often as possible. Early reviews by Peabody (1962) and Hamilton (1968) concluded that ERB is a consistent individual difference, and more recent studies have sustained this conclusion. Bachman and O'Malley (1984a) found ERB in attitudes to be highly stable over time. A study of trait ratings found ERB to be the major source of individual differences across raters (Van der Kloot, Kroonenberg, & Bakker, 1985). Race differences in ERB have also been found (Bachman & O'Malley, 1984a). There is little support, however, for a link between ERB and any traditional personality dimensions (Bonarius, 1971; Schneider, 1973).

Not all extreme responding represents a response bias in test-taking. There is substantial evidence that extreme responses are valid indicators of extreme opinions. Most convincing is the evidence that extreme test responses predict extreme behavior (Schuman & Presser, 1981; cf. Peabody, 1962).

Not to be confused with ERB is the so-called deviant response style (Berg, 1967), the tendency to make ratings as different as possible from the norm. Berg hypothesized links between this style and a wide range of behaviors and collected some relevant data. Its failure to gain credibility is summarized by Wiggins's (1973) statement that "There is good reason to question both the evidence for and the explanatory value of the deviation hypothesis" (p. 419).

Problems

ERB precludes the direct comparability of one subject's scores to another's: One cannot ordinarily distinguish whether an extreme rating indicates a strong opinion or a tendency to use the extremities of rating scales. A second problem is that ERB induces spurious correlations among otherwise unrelated constructs. Another source of problems is the interaction between ERB and social status variables such as gender (Hamilton, 1968), race (Bachman & O'Malley, 1984a), and education (Shulman, 1973).

Control

In some situations, ERB can be controlled by putting questions in multiple-choice format. ERB cannot be corrected simply by balancing the key because extremity operates in both directions. Reducing the number of response options to two does eliminate the problem but simultaneously reduces the sensitivity of the measure. Standardizing the within-subjects variance equates subjects on extremity but subject variances may contain content because they are often inextricably confounded with subject means. In measuring self-esteem, for example, most responses are on the positive portion of the rating scale, thus confounding high self-esteem and ERB (Bachman & O'Malley, 1984b).

Measurement

There are no standard instruments for assessing ERB as a response style. In some applications, the variance of a subject's ratings across an inventory has been used as an index

(e.g., Van der Kloot *et al.*, 1985). Of course, this is inappropriate if the key for each dimension is not balanced or if the means depart substantially from the scale midpoint. Note that if only one dimension is being assessed, it is difficult to distinguish any index of ERB from a measure of dimensional importance or salience for that topic.

Future Research Directions

Although the debates are less heated than in the 1960s, the issue of response bias in self-reports continues to be rather polarized through the 1990s. Some researchers, citing Block (1965) and Rorer (1965), believe that the fear of response biases is unwarranted. Others continue to balk at using instruments that have unbalanced keys or that correlate highly with some measure of SDR.

Consequently, the more recent disputes about SDR often give one a sense of *déjà vu*. In the well-being literature, for example, Carstensen and Cone (1983) alleged that the popular measures were invalid because of high correlations with Edwards's Social Desirability scale. McCrae (1986) responded that such correlations were due to the fact that social desirability scales measure substance, not style. A similar debate recently occurred over the high correlations found between SD and measures of depression (Linehan & Nielsen, 1981; Nevid, 1983). One possible resolution to such standoffs lies in the separation of SDR instruments of self-deception and impression management (e.g., Sabourin *et al.*, 1989). Advances on these issues will be possible only if researchers abandon the all-or-none positions taken in the past.

With regard to SDR, two critical issues need to be addressed by researchers in this area. One issue is the clarification of the link between adjustment and an exaggerated positivity. Do all well-adjusted persons exaggerate? Is it only the extreme scorers? Alternatively, is it only a subset of maladjusted individuals who defensively report being adjusted? (See Paulhus, Fridhandler, & Hayes, in press.)

The second pressing need is for increased specificity in the measurement of impression management (called for some time ago by Norman, 1963). Few would disagree that respondents bent on impression management tailor their self-reports to suit the audience. Yet the target audience on available scales is a vague notion of society at large (DeMaio, 1984). Among the most useful scales would be those targeted at such audiences as job interviewers, psychology experimenters, and college peers.⁷ Moreover, separate measures may be required to index faking good, faking bad, and faking mad (Furnham & Henderson, 1982; Winder *et al.*, 1975). Finally, more research is needed to clarify what impression management scores mean when completed under anonymous conditions.

The current situation is more stable with respect to acquiescence. In personality, the pro forma compromise has been to balance the keying of new instruments without necessarily conceding that acquiescence makes a difference. In survey research, some difficult issues remain (Schuman & Presser, 1981).

Extremity response bias is typically ignored by contemporary test developers, although evidence confirms that it is pervasive (Van der Kloot *et al.*, 1985). A number of researchers are actively pursuing questions about which item contexts (e.g., Eiser & van der Pligt, 1984; Romer, 1983) and emotional states (Paulhus, 1987) are conducive to extreme responding.

⁷The scales in the Hogan Personality Inventory (1986) represent various images of self presentation. These images, however, are cross-situationally stable because they derive from motives and eventually become automatic.

Some researchers dismiss the importance of response bias contamination by pointing to significant validity coefficients for their content scales. They must be reminded that modest validity coefficients leave much unexplained variance in the predictor, some of which may be response bias. Moreover, the impression management demand of most practical assessment situations makes it likely that at least some respondents are faking.

While I fully intend to end on a negative note, there is at least one reason to be optimistic about a future break in the deadlocks over the nature of response styles. That reason is the growing interest in the *process* of questionnaire responding (e.g., Cliff, 1977; DeBoeck, 1981; Jackson, 1986; Novakowska, 1970; Rogers, 1971, 1974; Schwarz, Strack, Muller, & Chassein, 1988). Much of this research exploits the techniques of modern cognitive psychology, for example, computer-controlled presentation of stimuli and measurement of reaction times (Holden, Fekken, & Jackson, 1985; Hsu, Santelli, & Hsu, 1989; Knowles, 1988; Paulhus & Levitt, 1987). It may be some time, however, before any benefits trickle down to test consumers.

Acknowledgments

Work on this chapter was supported by a grant from the Social Sciences Research Council of Canada. Among those who assisted by commenting on the chapter were Lew Goldberg, Stanley Presser, Peter Schmolok, Howard Schuman, and Jerry Wiggins.

Bibliography

- Allaman, J. D., Joyce, C. S., & Crandall, V. C. (1972). The antecedents of social desirability response tendencies of children and young adults. *Child Development*, *43*, 1135-1160.
- Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General*, *108*, 441-485.
- Bachman, J. G., & O'Malley, P. M. (1984a). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*, 491-509.
- Bachman, J. G., & O'Malley, P. M. (1984b). Black-white differences in self-esteem: Are they affected by response styles? *American Journal of Sociology*, *90*, 624-639.
- Becker, W. M. (1976). Biasing effect of respondents' identification on responses to a social desirability scale: A warning to researchers. *Psychological Reports*, *39*, 756-758.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, *76*, 186-204.
- Berg, I. A. (1967). The deviation hypothesis: A broad statement of its assumptions and postulates. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 146-190). Chicago: Aldine.
- Bernreuter, R. G. (1933). Validity of the personality inventory. *Personality Journal*, *11*, 383-386.
- Block, J. (1965). *The challenge of response sets*. New York: Appleton-Century-Crofts.
- Block, J. (1971). On further conjectures regarding acquiescence. *Psychological Bulletin*, *76*, 205-210.
- Bonarius, J. C. J. (1971). *Personal construct psychology and extreme response style*. Amsterdam: Swets and Zeitlinger.
- Borkenau, P., & Amelang, M. (1985). The control of social desirability in personality inventories: A study using the principal-factor deletion technique. *Journal of Research in Personality*, *19*, 44-53.
- Borkenau, P., & Ostendorf, F. (1989). Descriptive consistency and social desirability in self- and peer reports. *European Journal of Personality*, *3*, 31-45.
- Bramble, W. J., & Wiley, D. E. (1974). Estimating content-acquiescence correlation by covariance structure analysis. *Multivariate Behavioral Research*, *9*, 179-190.
- Brannigan, G. G. (1974). Comparison of yes-no and true-false forms of the children's social desirability scale. *Psychological Reports*, *34*, 898.
- Burish, T. G., & Houston, B. K. (1976). Construct validity of the Lie scale as a measure of defensiveness. *Journal of Clinical Psychology*, *32*, 310-314.
- Byrne, D. (1964). Repression-sensitization as a dimension of personality. In B. A. Maher (Ed.), *Progress in experimental personality research* (Vol. 1, pp. 169-220). New York: Academic Press.
- Carr, L. G. (1971). The Srole items and acquiescence. *American Sociological Review*, *36*, 287-293.
- Carstensen, L. L., & Cone, J. D. (1983). Social desirability and measurement of psychological well-being in elderly persons. *Journal of Gerontology*, *38*, 713-715.
- Cattell, R. B., Pierson, G., & Finkbeiner, C. (1976). Alignment of personality source trait factors from questionnaires and observer ratings: The theory of instrument-free patterns. *Multivariate Experimental Clinical Research*, *2*, 63-88.
- Clark, J. P., & Tift, L. L. (1966). Polygraph and interview validation of self-reported deviant behavior. *American Sociological Review*, *31*, 516-523.
- Cliff, N. (1977). Further study of cognitive processing models for inventory response. *Applied Psychological Measurement*, *1*, 41-49.
- Comrey, A. L. (1980). *Handbook of interpretations for the Comrey Personality scales*. San Diego: EdITS.
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*, 151-174.
- Crandall, V. C. (1966). Personality characteristics and social and achievement behaviors associated with children's social desirability response tendencies. *Journal of Personality and Social Psychology*, *4*, 477-486.
- Crandall, V. C., & Gozali, J. (1969). The social desirability responses of children of four religious-cultural groups. *Child Development*, *40*, 751-762.
- Crandall, V. C., Crandall, V. J., & Katkovsky, W. (1965). A children's social desirability questionnaire. *Journal of Consulting Psychology*, *29*, 27-36.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*, 475-494.
- Crowne, D. P. (1979). *The experimental study of personality*. Hillsdale, NJ: Erlbaum.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. New York: Wiley.
- Cunningham, M. R. (1989). Test-taking motivations and outcomes on a standardized measure of on-the-job integrity. *Journal of Business and Psychology*, *3*, 120-125.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (Eds.) (1972). *An MMPI handbook* (Vol. 1). Minneapolis: Univ. of Minnesota Press.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (Eds.) (1975). *An MMPI handbook* (Vol. 2). Minneapolis: Univ. of Minnesota Press.
- Damarin, F., & Messick, S. (1965). *Response styles as personality variables: A theoretical integration of multivariate research* (Res. Bull. No. 65-10). Princeton, NJ: Educational Testing Service.
- Davis, C., & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Measurement*, *49*, 311-320.
- Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 509-566). New York: Random House.
- DeBoeck, P. (1981). Individual differences in the validity of a cognitive processing model for responses to personality inventories. *Applied Psychological Measurement*, *5*, 481-492.
- DeLamater, J., & McKinney, K. (1982). Response-effects of question content. In W. Dijkstra & J. van der Zouwen (Eds.), *Response behaviour in the survey-interview*. New York: Academic Press.

- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257-282). New York: Russell Sage Foundation.
- Dillehay, R. C., & Jernigan, L. R. (1970). The biased questionnaire as an instrument of opinion change. *Journal of Personality and Social Psychology*, *15*, 144-150.
- 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 Press.
- Edwards, A. L. (1963). A factor analysis of experimental social desirability and response set scales. *Journal of Applied Psychology*, *47*, 308-316.
- Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories*. New York: Holt-Rinehart-Winston.
- Edwards, A. L., & Walsh, J. A. (1964). Response sets in standard and experimental personality scales. *American Educational Research Journal*, *1*, 52-61.
- Edwards, L. K., & Clark, C. L. (1987). Social desirability values and dispersions for the items in two SD scales: The MMPI SD scale and an experimental SD scale. *Psychological Reports*, *60*, 1083-1086.
- Edwards, L. K., Edwards, A. L., & Clark, C. L. (1988). Social desirability and the frequency of social-reinforcement scale. *Journal of Personality and Social Psychology*, *54*, 526-529.
- Eiser, J. R., & van der Pligt, J. (1984). Accentuation theory, polarization, and the judgment of attitude statements. In J. R. Eiser (Ed.), *Attitudinal judgment* (pp. 43-63). New York: Springer-Verlag.
- Eysenck, H. J., & Eysenck, S. B. G. (1964). *The manual of the Eysenck Personality Inventory*. London: Univ. of London Press.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *The Eysenck Personality Questionnaire manual*. London: Hodder & Stoughton.
- Faranda, J. A., Kaminski, J. A., & Giza, B. K. (1979). *An assessment of attitudes toward women with the bogus pipeline*. Paper presented at the meeting of the American Psychological Association, New York.
- Fisher, G. (1967). Normative and reliability data for the standard and the cross-validated Marlowe-Crowne Social Desirability Scale. *Psychological Reports*, *20*, 174.
- Folkman, S., Lazarus, R. S., Dunkel-Schetter, C., DeLongis, A., & Gruen, R. J. (1986). Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, *50*, 992-1003.
- Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences*, *7*, 385-400.
- Furnham, A., & Henderson, M. (1982). The good, the bad, and the mad: Response bias in self-report measures. *Personality and Individual Differences*, *3*, 311-320.
- Gocka, E. (1965). American Lake norms for 200 MMPI scales. Unpublished data cited in Dahlstrom, Welsh, & Dahlstrom (1975).
- Goldberg, L. R. (1972). Student personality characteristics and optimal college learning conditions: An extensive search for trait-by-treatment interaction effects. *Instructional Science*, *1*, 153-210.
- Goode, W. J., & Hart, P. K. (1952). *Methods in social science*. New York: McGraw-Hill.
- Gough, H. G. (1947). Simulated patterns on the Minnesota Multiphasic Personality Inventory. *Journal of Abnormal and Social Psychology*, *42*, 215-225.
- Gough, H. G. (1952). On making a good impression. *Journal of Educational Research*, *46*, 33-42.
- Gough, H. G. (1957). *California Psychological Inventory manual*. Palo Alto, CA: Consulting Psychologists Press.
- Gough, H. G. (1987). *California Psychological Inventory administrator's guide*. Palo Alto, CA: Consulting Psychologists Press.
- Gough, H. G., & Heilbrun, A. B. (1965). *The adjective checklist manual*. Palo Alto, CA: Consulting Psychologists Press.

- Gough, H. G., & Heilbrun, A. B. (1980). *The adjective checklist manual*. Palo Alto, CA: Consulting Psychologists Press.
- Gove, W. R., & Geerken, M. (1977). Response bias in surveys of mental health: An empirical investigation. *American Journal of Sociology*, *82*, 1289-1317.
- Greenberg, B. C., Abdula, A. L., Simmons, W. L., & Horvitz, D. G. (1969). The unrelated question in randomized response model, theoretical framework. *Journal of the American Statistical Association*, *64*, 520-539.
- Greene, R. L. (1980). *The MMPI: An interpretive manual*. New York: Grune & Stratton.
- Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*, 192-203.
- Hartshorne, H., & May, M. A. (1928). *Studies in deceit*. New York: Macmillan.
- Hathaway, S. R., & McKinley, J. C. (1951). *The MMPI manual*. New York: Psychological Corporation.
- Heilbrun, A. B. (1961). The psychological significance of the MMPI K scale in a normal population. *Journal of Consulting Psychology*, *25*, 486-491.
- Heilbrun, A. B. (1963). Revision of the MMPI K correction for improved detection of maladjustment in a normal college population. *Journal of Consulting Psychology*, *27*, 161-165.
- Helmes, E., & Holden, R. R. (1986). Response styles and faking on the Basic Personality Inventory. *Journal of Consulting and Clinical Psychology*, *54*, 853-859.
- Hogan, R. (1986). *Hogan Personality Inventory*. Minneapolis, MN: National Computer Systems.
- Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology*, *74*, 273-279.
- Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality*, *23*, 180-191.
- Holden, R. R., Fekken, G. C., & Jackson, D. N. (1985). Structured personality test item characteristics and validity. *Journal of Research in Personality*, *19*, 386-394.
- Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, *53*, 278-295.
- Ihlevich, D., & Gleser, G. C. (1986). *Defense mechanisms: Their classification, correlates, and measurement with the Defense Mechanisms Inventory*. Owosso, MI: DMI Associates.
- Jackson, D. N. (1967). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Research Psychologists Press.
- Jackson, D. N. (1986). The process of responding in personality assessment. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire: Current issues in theory and measurement* (pp. 123-142). New York: Springer-Verlag.
- Jackson, D. N., & Helmes, E. (1979). Personality structure and the circumplex. *Journal of Personality and Social Psychology*, *37*, 2278-2285.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*, 243-252.
- Jackson, D. N., & Messick, S. (1962). Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition* (pp. 129-155). New York: Wiley.
- Jacobson, L. I., Berger, S. E., & Millham, J. (1970). Individual differences in cheating during a temptation period when confronting failure. *Journal of Personality and Social Psychology*, *15*, 48-56.
- Jacobson, L. I., Brown, R. F., & Ariza, M. J. (1983). A revised multidimensional social desirability inventory. *Bulletin of the Psychonomic Society*, *21*, 391-392.
- Jacobson, L. I., Kellogg, R. W., Cauce, A. M., & Slavin, R. S. (1977). A multidimensional social desirability inventory. *Bulletin of the Psychonomic Society*, *9*, 109-110.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*, 349-364.
- Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin*, *85*, 555-586.

- Karson, S., & O'Dell, J. W. (1976). *A guide to the clinical use of the 16PF*. Champaign, IL: Institute for Personality and Ability Testing.
- Katkin, E. S. (1964). The Marlowe-Crowne Social Desirability Scale: Independent of psychopathology? *Psychological Reports*, *15*, 703-706.
- Kiecolt-Glaser, J., & Murray, J. A. (1980). Social desirability bias in self-monitoring data. *Journal of Behavioral Assessment*, *2*, 239-247.
- Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, *55*, 312-320.
- Knudson, D. D., Pope, H., & Irish, D. P. (1967). Response differences to questions on sexual standards. *Public Opinion Quarterly*, *31*, 290-297.
- Kozma, A., & Stones, M. J. (1988). Social desirability in measures of subjective well-being: Age comparisons. *Social Indicators Research*, *20*, 1-14.
- Krebs, D., & Schuessler, K. (Eds.) (1987). *Soziale Empfindungen: Ein interkultureller Skalenvergleich bei Deutschen und Amerikanern*. New York/Frankfurt: Campus Verlag.
- Lanning, K. (1989). Detection of invalid response patterns on the California Psychological Inventory. *Applied Psychological Measurement*, *13*, 45-56.
- Lautenschlager, G. J. (1986). Within-subject measures for the assessment of individual differences in faking. *Educational and Psychological Measurement*, *46*, 309-316.
- Lautenschlager, G. J., & Flaherty, V. L. (in press). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*.
- Leak, G. K., & Fish, S. (1989). Religious orientation, impression management, and self-deception. Toward a clarification of the link between religiosity and social desirability. *Journal for Scientific Study of Religion*, *28*, 355-359.
- Lenski, G. E., & Leggett, J. C. (1960). Caste, class and deference in the research interview. *American Journal of Sociology*, *65*, 463-467.
- Lentz, T. F. (1938). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin*, *35*, 659.
- Linden, W., Paulhus, D. L., & Dobson, K. S. (1986). Effects of response styles on the report of psychological and somatic distress. *Journal of Consulting and Clinical Psychology*, *54*, 309-313.
- Linehan, M. M., & Nielsen, S. L. (1981). Assessment of suicide ideation and parasuicide: Hopelessness and social desirability. *Journal of Consulting and Clinical Psychology*, *49*, 773-775.
- Lockard, J. S., & Paulhus, D. L. (Eds.) (1988). *Self-deception: An adaptive mechanism?* New York: Prentice-Hall.
- Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, *74*, 72-80.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology*, *43*, 293-303.
- McCrae, R. R. (1986). Well-being scales do not measure social desirability. *Journal of Gerontology*, *41*, 390-392.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*, 882-888.
- McCrae, R. R., Costa, P. T., Dahlstrom, W. G., Barefoot, J. C., Siegler, I. C., & Williams, R. B. (1989). A caution on the use of the MMPI K-correction in research on psychosomatic medicine. *Psychosomatic Medicine*, *51*, 58-65.
- McKinley, J. C., Hathaway, S. R., & Meehl, P. E. (1948). The Minnesota Multiphasic Personality Inventory. IV. The K scale. *Journal of Consulting Psychology*, *12*, 20-31.
- Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*, 525-564.
- Mellor, S., Conroy, L., & Masteller, B. K. (1986). Comparative trait analysis of long-term recovering alcoholics. *Psychological Reports*, *58*, 411-418.
- Messick, S. (1967). The psychology of acquiescence: An interpretation of research evidence. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 115-145). Chicago: Aldine.
- Millham, J. (1974). Two components of need for approval score and their relationship to cheating following success and failure. *Journal of Research in Personality*, *8*, 378-392.

- Millham, J., & Jacobson, L. I. (1978). The need for approval. In H. London & J. E. Exner (Eds.), *Dimensions of personality* (pp. 365-390). New York: Wiley.
- Millham, J., & Kellogg, R. W. (1980). Need for social approval: Impression management or self-deception? *Journal of Research in Personality*, *14*, 445-457.
- Montag, I., & Comrey, A. L. (1982). Personality construct similarity in Israel and the United States. *Applied Psychological Measurement*, *6*, 61-67.
- Morf, M. E., & Jackson, D. N. (1972). An analysis of two response styles: True responding and item endorsement. *Educational and Psychological Measurement*, *32*, 329-353.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*, 263-280.
- Nevid, J. S. (1983). Hopelessness, social desirability and construct validity. *Journal of Consulting and Clinical Psychology*, *51*, 139-140.
- Norman, W. T. (1963). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, *47*, 225-241.
- Norman, W. T. (1967). On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin*, *67*, 273-293.
- Norman, W. T. (1969). "To see ourselves as others see us!": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research*, *4*, 417-443.
- Norman, W. T. (1990). On separating substantive, stylistic, and evaluative components in personality measurements: A cross-national comparison. *Journal of Personality and Social Psychology*.
- Novakowska, M. (1970). A model of answering to a questionnaire item. *Acta Psychologica*, *34*, 420-439.
- Paulhus, D. L. (1981). Control of social desirability in personality inventories: Principal-factor deletion. *Journal of Research in Personality*, *15*, 383-388.
- Paulhus, D. L. (1982). Individual differences, self-presentation, and cognitive dissonance: Their concurrent operation in forced compliance. *Journal of Personality and Social Psychology*, *43*, 838-852.
- 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 questionnaire* (pp. 143-165). New York: Springer-Verlag.
- Paulhus, D. L. (1987). *Emotional arousal and evaluative extremity: A dynamic complexity analysis*. Manuscript submitted for publication.
- Paulhus, D. L. (1988). *Assessing self-deception and impression management in self-reports: The Balanced Inventory of Desirable Responding*. Unpublished manual, University of British Columbia, Vancouver, Canada.
- Paulhus, D. L., Fridhandler, B., & Hayes, S. (in press). Psychological defense. In S. R. Briggs, R. Hogan, & W. Jones (Eds.), *Handbook of personality psychology*. San Diego, CA: Academic Press.
- Paulhus, D. L., & Levitt, K. (1987). Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology*, *52*, 245-259.
- Paulhus, D. L., & Reid, D. B. (in press). Attribution and denial in socially desirable responding. *Journal of Personality and Social Psychology*.
- Paulhus, D. L., Graf, P., & Van Selst, M. (1989). Attentional load increases the positivity of self-presentation. *Social Cognition*, *7*, 389-400.
- Peabody, D. (1962). Two components in bipolar scales: Direction and extremeness. *Psychological Review*, *69*, 65-73.
- Peabody, D. (1966). Authoritarianism scales and response bias. *Psychological Bulletin*, *65*, 11-23.
- Phillips, D. L., & Clancy, K. J. (1972). Some effects of social desirability in survey studies. *American Journal of Sociology*, *77*, 921-940.
- Quigley-Fernandez, B., & Tedeschi, J. T. (1978). The bogus pipeline as lie-detector: Two validity studies. *Journal of Personality and Social Psychology*, *36*, 247-256.

- Quinn, B. A. (1989). *Religiousness and psychological well-being: An empirical investigation*. Unpublished dissertation, Wayne State Univ., Detroit.
- Ray, J. J. (1983). Reviving the problem of acquiescent response bias. *Journal of Social Psychology*, *121*, 81-96.
- Reis, H. A. (1966). The MMPI K scale as a predictor of prognosis. *Journal of Clinical Psychology*, *22*, 212-213.
- Riggio, R. E., Salinas, C., & Tucker, J. (1988). Personality and deception ability. *Personality and Individual Differences*, *9*, 189-191.
- Rogers, T. B. (1971). The process of responding to personality items: Some issues, a theory and some research. *Multivariate Behavioral Research Monographs*, 6 Whole No. 2.
- Rogers, T. B. (1974). An analysis of two central stages underlying responding to personality items: The self-referent decision and response selection. *Journal of Research in Personality*, *8*, 128-138.
- Romer, D. (1983). Effects of own attitude on polarization of judgment. *Journal of Personality and Social Psychology*, *44*, 273-284.
- Rorer, L. G. (1965). The great response style myth. *Psychological Bulletin*, *63*, 129-156.
- Rorer, L. G., & Goldberg, L. R. (1965). Acquiescence and the vanishing variance component. *Journal of Applied Psychology*, *49*, 422-430.
- Roth, D. L., Harris, R. N., & Snyder, C. R. (1988). An individual differences measure of attributive and repudiative tactics of favorable self-presentation. *Journal of Social and Clinical Psychology*, *6*, 159-170.
- Roth, D. L., Snyder, C. R., & Pace, L. M. (1986). Dimensions of favorable self-presentation. *Journal of Personality and Social Psychology*, *51*, 867-874.
- Ruch, F. L. (1942). A technique for detecting attempts to fake performance on the self-inventory type of personality test. In J. F. Dashiell (Ed.), *Studies in personality*. New York: McGraw-Hill.
- Ruch, F. L., & Ruch, W. W. (1967). The K factor as a (validity) suppressor variable in predicting success in selling. *Journal of Applied Psychology*, *51*, 201-204.
- Sabourin, S., Bourgeois, L., Gendreau, P., & Morval, M. (1989). Self-deception, impression management, and consumer satisfaction with mental health treatment. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *1*, 126-129.
- 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: Plenum.
- Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology*, *47*, 213-215.
- Sackett, P. R., & Harris, M. M. (1984). Honesty testing for personnel selection: A review and critique. *Personnel Psychology*, *37*, 221-245.
- Sackett, P. R., Burrell, L. R., & Callahan, C. (1989). *Integrity testing for personnel selection: An update*. Unpublished manuscript, University of Minnesota, Minneapolis.
- Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*, 219-247.
- Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, *79*, 294-309.
- Schriesheim, C. A. (1979). Social desirability and leader effectiveness. *Journal of Social Psychology*, *108*, 89-94.
- Schuessler, K., Hittle, D., & Cardascia, J. (1978). Measuring responding desirably with attitude-opinion items. *Social Psychology*, *41*, 224-235.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Schwarz, N., Strack, F., Muller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, *6*, 107-117.
- Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. *Personnel Psychology*, *39*, 791-809.

- Shulman, A. (1973). A comparison of two scales on extremity response bias. *Public Opinion Quarterly*, *37*, 407-412.
- Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, *18*, 247-255.
- Sigall, H., & Page, R. (1972). Reducing attenuation in the expression of interpersonal affect via the bogus pipeline. *Sociometry*, *35*, 629-642.
- Staub, E., & Sherik, L. (1970). Need for approval, children's sharing behavior and reciprocity in sharing. *Child Development*, *41*, 243-252.
- Strickland, B. R. (1977). Approval motivation. In T. Blass (Ed.), *Personality variables in social behavior* (pp. 315-356). Hillsdale, NJ: Erlbaum.
- Strosahl, K. D., Linehan, M. M., & Chiles, J. A. (1984). Will the real social desirability scale please stand up?: Hopelessness, depression, social desirability, and the prediction of suicidal behavior. *Journal of Consulting and Clinical Psychology*, *52*, 449-457.
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago: Aldine.
- Sutherland, B. V., & Spilka, B. (1964). Social desirability, item-response time, and item significance. *Journal of Consulting Psychology*, *28*, 447-451.
- Sweetland, A., & Quay, H. A. (1953). A note on the K scale of the MMPI. *Journal of Consulting Psychology*, *17*, 314-316.
- Swenson, W. M., Pearson, J. S., & Osbourne, D. (1973). *An MMPI sourcebook: Basic item, scale, and pattern data on 50,000 medical patients*. Minneapolis: Univ. of Minnesota Press.
- Tanaka-Matsumi, J., & Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *Journal of Consulting and Clinical Psychology*, *54*, 328-333.
- Taylor, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology*, *48*, 285-290.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social-psychological perspective on mental health. *Psychological Bulletin*, *103*, 193-210.
- Tellegen, A. (1982). *Brief manual for the Differential Personality Inventory*. Unpublished manuscript, University of Minnesota, Minneapolis.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, *103*, 299-314.
- Trott, D. M., & Jackson, D. N. (1967). An experimental analysis of acquiescence. *Journal of Experimental Research in Personality*, *2*, 278-288.
- Van der Kloot, W. A., Kroonenberg, P. M., & Bakker, D. (1985). Implicit theories of personality: Further evidence of extreme response style. *Multivariate Behavioral Research*, *20*, 369-387.
- Vernon, P. E. (1934). The attitude of the subject in personality testing. *Journal of Applied Psychology*, *18*, 165-177.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63-69.
- Webster, H. (1958). Correcting personality scales for response sets or suppression effects. *Psychological Bulletin*, *55*, 62-64.
- Weinberger, D. A. (in press). The construct validity of the repressive coping style. In J. L. Singer (Ed.), *Repression and dissociation: Defense mechanisms and personality styles*. Chicago: Univ. 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.
- Weiss, C. H. (1968). Validity of welfare mothers' interview responses. *Public Opinion Quarterly*, *32*, 622-633.
- Wiggins, J. S. (1964). Convergences among stylistic response measures from objective personality tests. *Educational and Psychological Measurement*, *24*, 551-562.
- Wiggins, J. S. (1968). Personality structure. *Annual Review of Psychology*, *19*, 293-350.
- Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading MA: Addison-Wesley.

- Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, *37*, 395-412.
- Winder, P., O'Dell, J. W., & Karson, S. (1975). New motivational distortion scales for the 16 PF. *Journal of Personality Assessment*, *39*, 532-537.
- Winkler, J. D., Kanouse, D. E., & Ware, J. E., Jr. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, *67*, 555-561.
- Winters, K. C., & Neale, J. M. (1985). Mania and low self-esteem. *Journal of Abnormal Psychology*, *94*, 282-290.
- Wiseman, F. (1972). Methodological bias in public opinion surveys. *Public Opinion Quarterly*, *36*, 105-108.
- Wooten, A. J. (1984). Effectiveness of the K correction in the detection of psychopathology and its impact on profile height and configuration among young adult men. *Journal of Consulting and Clinical Psychology*, *52*, 468-473.
- Wright, J. D. (1975). Does acquiescence bias the "index of political efficacy"? *Public Opinion Quarterly*, *39*, 219-226.
- Yonge, G. D. (1966). Certain consequences of applying the K factor to MMPI scores. *Educational and Psychological Measurement*, *26*, 887-893.
- Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizations: A reconception. *Academy of Management Review*, *12*, 250-264.