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 Sackheim 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 Ilievich and Gieser’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 nonexistent 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**


**Comments**

The predecessors of these measures, the Self-Deception and Other-Deception Questionnaires, were first described in Sackheim 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.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOT TRUE</td>
<td>SOMEWHAT</td>
<td>TRUE</td>
<td>VERY TRUE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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
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 Jack-
son's (1967) Desirability scale.

*Discriminant*

According to Schuessler et al. (1978), discriminant validity is supported by low intercor-
rrelations 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*


*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 presenta-
tion: 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. Nonexistent 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. Allman 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

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.
___ 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., Lenz, 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, 1984a; 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).

---

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

\(^6\) Sample 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 (Linchan & 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. 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 Schmolk, Howard Schuman, and Jerry Wiggins.

Bibliography


2. Measurement and Control of Response Bias


