CHAPTER 13

The Self-Report Method

Delroy L. Paulhus
Simine Vazire

In R.W. Robins, R.C.Fraley, & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp.224-239). New York: Guilford.

If you want to know what Waldo is like, why not just ask him? Such is the commonsense logic behind the self-report method of personality assessment. It remains the field's most commonly used mode of assessment—by far (see Robins, Tracy, & Sherman, Chapter 37, this volume). Despite its popularity and demonstrated utility, the self-report method has been a frequent target of criticism from the early days of psychological assessment (Allport, 1927) right up to the present (Dunning, Heath, & Suls, 2005).

The psychological processes underlying an act of self-reporting are now understood to be exceedingly complex (e.g., Hogan & Nicholson, 1988; Johnson, 2004; Schwarz, 1999; Tourangeau, Rips, & Rasinksi, 2001). Examination of these processes requires burrowing deep into the affective and cognitive substrates of personality. Among the challenging issues are the role of motives in self-perception (Robins & John, 1997), the applicability of performative models (Johnson, 2004), the effectiveness of introspection (Wilson, 2002), the degree of automaticity (Mills & Hogan, 1978; Paulhus & Levitt, 1986), and the meaning of nonresponding (Tourangeau, 2004).

The goal of this chapter is more limited: to provide a brief guide to nonexpert researchers interested in using the self-report method to assess personality. We begin by delineating three categories of self-reports. We then review the advantages and the disadvantages of the self-report method. Next, we examine the convergence of self-reports with other methods of assessing personality. Finally, we provide a practical guide to choosing a self-report instrument.

Types of Self-Reports

Variants of the self-report method are numerous and could be organized in a number of ways. We restrict ourselves to cases in which re-

spondents are aware that they are reporting on their personalities. Thereby ruled out are such methods as projective tests, handwriting analysis, conditional reasoning, and nonverbal coding techniques. As a result, we are left with three broad categories.

Direct Self-Ratings

In the simplest form of the self-report method, people are asked to report directly on their own personalities. In the case of a *global self-rating*, the respondent is furnished with a face-valid label of the construct and asked to give a summary self-appraisal. For some constructs, a single global rating can be surprisingly valid (Burisch, 1983). A recent example is the Single-Item Self-Esteem Scale (Robins, Hendin, & Trzesniewski, 2001). Other successful examples are the Five-Item Personality Inventory (FIPI) (Gosling, Rentfrow, & Swann, 2003) and the Single-Item Measures of Personality (SIMP)(Woods & Hampson, 2005), which capture each of the Big Five factors with a single item.

The authors of these single-item measures took pains to ensure that the item supplied a clear description of the attribute being assessed. In short, they sought high *face validity.* Of course, the validity of a clear item depends on the respondent's willingness and ability to provide that information (see below). The point here is that earnest attempts to disclose one's personality are facilitated by a clear question.

As a general rule, however, single-item assessment is not recommended because its reliability is usually lower than that of multi-item composites.[1] The availability of multiple items also makes it easier to control for certain response styles such as acquiescence and extremity (see below). Because a variety of items are administered to assess each construct, it may be less obvious what the test is designed to assess. Nonetheless, respondents usually try to make sense of the test and their interpretation gets more confident as more items are presented (Knowles & Condon, 1999). To help clarify the construct to the respondent, items can be grouped and labeled (Goldberg, 1992).

Some constructs (e.g., openness to experience, ego-resiliency) are too complex to be directly rated and, even with multiple items, may never crystallize in the mind of the respondent. Nonetheless, the aggregation of items may yield a scientifically meaningful score.

To summarize, the key features in the direct-rating approach—especially the global self-rating—are clarity and simplicity. In many cases, a direct request for personality information will yield the most valid assessment.

Indirect Self-Reports

Like direct self-ratings, *indirect self-reports* pose questions about the respondent's personality. The primary difference is that indirect self-reports usually obscure the construct being measured: Respondents may even be intentionally misled about the purpose of the test.[2] For example, the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1981) asks respondents about their competence, their leadership ability, their storytelling ability, and their physical attractiveness. In truth, the NPI is not actually targeting any of those traits. Instead, a respondent accumulates a high narcissism score by repeatedly choosing the grandiose option across that whole range of superlative qualities.

Note that a corresponding self-rating measure of narcissism would simply ask for degree of agreement with the item "I am narcissistic." Because narcissists are defensive and deluded, however, a direct measure would be pointless. Thus, the choice of the indirect approach was theoretically driven.

A similar logic applies to several measures of socially desirable responding (e.g., the Marlowe-Crowne scale or the Balanced Inventory of Desirable Responding). For example, one item on the latter instrument reads "My first impressions about people are always right" (Paulhus, 1991). Yet the test aims to measure not interpersonal perspicuity, but the tendency to ascribe desirable (yet highly unlikely) qualities to oneself. Rather than the specific content, a respondent's score is based on the total number of desirable but unlikely qualities claimed. Indirect approaches have also been tried in the measurement of defense mechanisms via self-report (but see Davidson & MacGregor, 1998).

Another type of indirect test involves the use of *subtle items.* Here, the purpose of the item is obscure. The 49-item version of the MMPI Alcoholism Scale, for example, is composed entirely of subtle items (e.g., "I have had periods when I carried out activities without knowing later what I had been doing" and "My hardest battles are with myself"). They do not directly mention, but are predictive of an alcoholic pre-

disposition. The subtle approach has been subjected to substantial research, most of it indicating that subtle items are actually less valid than more obvious items (e.g., Holden, Fekken, & Jackson, 1985; Osberg, 1999).

In sum, indirect tests are designed to minimize face validity: The assessment of personality is based on the researcher's interpretation of the answers, not on the content of the respondent's intended message. The major advantage claimed for such measures is resistance to faking: If the respondent has no idea what the administrator is assessing, then faking is precluded.

Nonetheless, respondents do make inferences about what is being assessed. With an indirect test, however, the respondent's interpretation is likely to differ from the researcher's. Moreover, different respondents tend to make different inferences. For example, high scorers on anxiety scales assume that the test measures emotional honesty, whereas low scorers assume that it measures mental illness. High scorers on integrity tests assume that the assessor will hire people who admit no past misbehavior; low scorers take a more cynical stance in assuming that the assessor will take admissions of misbehavior as an indicator of honesty (Cunningham, Wong, & Barbee, 1994). The fact that such instruments show predictive validity may result from the finding that the differential interpretation is diagnostic by itself.

Note that the distinction between direct and indirect self-reports involves a continuum, rather than a strict dichotomy. Many measures are neither completely direct nor completely indirect. On most Big Five inventories, for example, the items are mixed up or rotated and, accordingly, only partly transparent. With enough repetition in the items, the respondent may eventually see the pattern.

Open-Ended Self-Descriptions

In this third general category, self-reports are derived from participants' free descriptions of their own personalities. Unlike structured measures (such as those described above), this open-ended form of self-report allows respondents to use any constructs they wish in describing themselves. The researcher may request a focus on certain trait domains, or be as loose as possible with an instruction such as "Describe your personality in the space provided." For example, the Twenty Statements Test (TST; Kuhn & McPartland, 1954) asks respondents to complete 20 sentences, each beginning with the phrase "I am."

For such self-descriptions to be quantified, they must be content coded. A systematic coding system must be developed and its meaningfulness confirmed by establishing good interrater reliabilities (see Woike, Chapter 17, this volume, for an example). Although the codings involve some degree of subjectivity, high reliabilities can often be established. For example, narcissism might be assessed by coding for (1) self-focus on positive as opposed to negative qualities and (2) implicit derogation of others. Note that the coding process is often protracted and labor-intensive, and there is no guarantee that a reliable and valid coding scheme can ever be developed.

Objective elements can also be coded from free descriptions. For example, narcissism might be measured by the sheer volume of self-description or the proportion of pronouns that are in the first person singular (*I* or *me*). Trait-relevant words can be indexed using the computer program Linguistic Inquiry and Word Count (LIWC) developed by Pennebaker, Francis, and Booth (2001). Note that some subjectivity is involved in mapping language use onto traits: The researcher must rationally designate which categories of words are indicators of which traits (see Vazire, Gosling, Dickey, & Schapiro, Chapter 11, this volume).

In sum, all three categories involve self-reports in the sense that respondents are aware that they are being asked about their personalities. Although the methods vary with respect to the assumption that respondents are capable of reporting their own personalities, and with respect to the amount of structure provided, they all share the assumption that self-reports can tell us a great deal about personality. Except where otherwise indicated, the issues discussed in the rest of this chapter apply to all three types of self-reports.

Advantages of Self-Reports

> *The best criterion for a target's personality is his or her self-rating. . . . Otherwise, the whole enterprise of personality assessment seriously needs to rethink itself.*
>
> —Reviewer A

Like Reviewer A, many researchers take for granted that self-reports are the ultimate measure of personality. Although alternative methods have an equally long history, self-reports remain the most popular choice. Their popularity appears to be based on a number of persuasive advantages: These include easy interpretability, richness of information, motivation to report, causal force, and sheer practicality. A substantial body of research supports all of these claims (Lucas & Baird, 2006; Swann, Chang-Schneider, & McClarty, in press).

Interpretability. Self-reports are communicated in the language common to the assessor and the respondent. Although there may be some variation within a culture, the assessor's request to a literate adult for a rating of, say, anxiety, can reasonably be assumed to be understood. This feature is one shared with informant reports. But compare that verbal equivalence to the case of behavioral measures such as galvanic skin response, heart rate, and blood pressure. Such behavioral measures are always subject to multiple interpretations (Wiggins, 1973). Life data, although often objectively scored, entail even more ambiguity. Variables such as socioeconomic status, educational level, and longevity are most certainly multiply determined and impossible to equate with a single psychological construct (Loevinger, 1957).

Information Richness

The notion that people are the best-qualified witnesses to their own personalities is supported by the indisputable fact that no one else has access to more information. First, the self has an the opportunity to observe a wide range of behaviors covering a wide swath of time. These include behaviors that are typically performed in private—for example, masturbation, academic cheating, napping, and singing in the shower. In short, people have access to a great *quantity* and *breadth* of information. Second, the self has access to intrapsychic information—thoughts, feelings, and sensations that are unavailable to others (Robins, Norem, & Cheek, 1999). In this sense, people possess a better *quality* of information about themselves. For example, an observer witnessing the theft of a candy bar has no access to the underlying motivation: The thief may have been motivated by thrill seeking, hunger, or the desire to make a political statement against corporations. In general, the self is better able to contextualize when reporting on personality-relevant information and, therefore, better able to provide a valid report.

Even when the same information is available to self and others, people are more likely to remember information that is self-relevant (Rogers, Kuiper, & Kirker, 1977). It is not surprising, then, that people's self-schemas are especially well developed. Although not necessarily improving the accuracy of self-perceptions, a well-developed schema certainly quickens the speed of trait-relevant information retrieval (Fekken & Holden, 1992).

Motivation to Report

Another advantage of the self-report method is that no one is more fascinated by the target of assessment than is the assessor. Up to a point, then, people are usually pleased to talk about themselves. Of course, the motivation to reflect on the self varies across individuals (Trapnell & Campbell, 1999).

Self-preoccupation may also lead people to answer more diligently when completing self-reports. Whereas ratings of others may be done carelessly or superficially, people tend to put in more time and effort when reporting on their own personalities. Greater validity should ensue. This argument applies all the more when respondents are completing questionnaires in order to get private feedback on their personalities.

Causal Force

Another advantage of the self-report method is that it engages the respondent's identity (Hogan & Smither, 2001). Accurate or not, self-perceptions have a strong influence on how people interact with the world. They affect behavior (Ickes, Snyder, & Garcia, 1997), self-presentation to others (Vazire & Gosling, 2004), and expectations about how one will be seen by others.

Although not synonymous with personality, identity is unquestionably a central aspect of personality (McAdams, 2000). A person's identity reflects the phenomenological experience of his or her personality: "What it feels like to be me." If someone thinks of him- or herself as neurotic, whether or not the objective evidence supports the person's self-view, that perception constitutes one aspect of his or her personality.

Although identifying oneself as shy may not be equivalent to being genetically shy, it does influence one's behavior (Cheek & Watson, 1989). In some cases, simply being asked to characterize oneself can influence one's future behavior (Greenwald, Carnot, Beach, & Young, 1987). In sum, the causal force of self-perceptions make them important to personality assessment in a rather different fashion from other indicators—reputation, for example (Hogan & Smither, 2001).

Practicality

Finally, a singular advantage of self-reports is their extraordinary practicality. They are both efficient and inexpensive. They require only the cooperation of the target person; in contrast, the collection of informant ratings, behavior assessment, or life data all require the involvement of less available parties. Self-reports are efficient because they can be administered in mass testing sessions (as opposed to one-on-one interviews, for example). Hundreds of variables can easily be collected in one sitting.

Self-reports also tend to be the least expensive data source. Although researchers may need to provide compensation for more effortful, intensive methods (e.g., diary studies), questionnaire administrations seldom require more incentive than the opportunity to express oneself. When an incentive is needed, provision of personality feedback will often suffice; for students, extra course credit will do. Indeed, self-reporters often volunteer and will sometimes pay to be assessed. Expenses for study materials seldom go beyond those for photocopying, and even these may be eliminated if the items are administered via the Internet. Data entry costs can be minimized by collecting responses on machine-scorable sheets.

In the case of some constructs, self-report is the only appropriate method (Ozer & Reise, 1994). For example, researchers interested in self-efficacy, a construct that is by nature a self-perception, must obtain self-reports. Self-esteem researchers often argue that methods other than self-report are simply inadequate (Robins et al., 2001). Other personality-related concepts best measured via self-report include well-being (Diener, Sandvik, Pavot, & Gallagher, 1991), values (Trapnell, 2006), personal projects (Little, 1998), and life goals (Roberts, O'Donnell, & Robins, 2004).

Beyond their virtues, self-reports are often a necessity: They may be the only available method. Survey research, for example, is entirely self-reported (Tourangeau, 2004). Internet studies of personality rely on self-reports: Responses are completed anonymously and, therefore, are not easily linked to other corroborative measures such as behavior or informant reports (Gosling, Vazire, Srivastava, & John, 2004).

Disadvantages of Self-Reports

Self-reports suffer from many of the same measurement artifacts as other assessment methods. These include anchoring effects, primacy and recency effects, time pressure, and consistency motivation. Such issues are beyond the scope of this chapter; instead, we focus on several problems that are unique to the self-report method.

An overarching issue is the credibility of self-reports. Why should we trust what people say about themselves? Clearly, accuracy is not the only motive shaping self-perceptions (Sedikides & Strube, 1995). Among the other powerful motives are consistency seeking, self-enhancement, and self-presentation (Robins and John, 1997; Swann et al., in press). Even when respondents are doing their best to be forthright and insightful, their self-reports are subject to various sources of inaccuracy. Of special interest, as discussed below, are limitations such as self-deception and memory.

Self-reports in the context of face-to-face interviews raise a host of other problems such as effects of self-consciousness, rapport, transference, and modeling. Unique issues are raised by computerized testing (Butcher, 2003) and Internet surveys (Gosling et al., 2004). Such issues go beyond the scope of our review, and we restrict our discussion to self-reports in the form of pencil-and-paper questionnaires.

Classic Response Sets and Styles

Some people show a tendency to respond to questions in a manner that, although systematic, interferes with the validity of the response (for a review, see Paulhus, 1991). Well-known examples are socially desirable responding (SDR), acquiescent responding (AR), and extreme responding (ER). When specific to the situation, these tendencies are termed *response*

sets; when consistent across time and assessment context, they are termed *response styles*.

Self-Presentation

We use the term *self-presentation* to subsume all self-report propensities of an evaluative nature. It includes self-aware forms—*impression management*—as well as unconscious forms—*self-deception*. Impression management includes such variants as exaggeration, faking, and lying, whereas self-deception includes variants such as self-favoring bias, self-enhancement, defensiveness, and denial. Self-presentation can be negative, as in *malingering* (Morey & Lanier, 1998), but the more common concern in personality research has been with positive biases such as SDR (Paulhus, 2002).

A temporary SDR set can be induced when a situational press compels the individual to give an overly positive self-description. For example, a job applicant may have such motivation to land a specific job that she exaggerates her credentials only on that one assessment. More or less random events (e.g., a recent job loss due to downsizing) can create a press for self-presentation that is independent of personality and ability and is, perforce, unpredictable. This response set should pervade all the variables in the specific assessment context but should not generalize to other contexts.

When stable across time and different questionnaires, this trait-like form of SDR is considered to be a response style (Edwards, 1957). Instruments are available to capture several variants of this concept (e.g., impression management, self-deceptive enhancement, self-deceptive denial, malingering) . Eight popular instruments were reviewed in detail by Paulhus (1991).

Originally, response styles were assumed to be limited to the questionnaire context. The consistency of styles across time and settings, however, implicates underlying personality traits (see below). Nonetheless, unless otherwise noted, our subsequent discussion and recommendations apply to SDR sets as well as styles.

The nature of concern about SDR differs somewhat between survey and personality researchers. Survey researchers would worry about an overall bias even if all respondents engaged in SD to the same degree. In personality research, however, the primary concern is that respondents may *differ* in their tendency to engage in SDR because it creates a confounding between SDR and personality content scales.

CONTROL OF SD EFFECTS

The litany of methods aimed at controlling SD contamination can be organized into three categories: (1) rational techniques, (2) demand reduction, and (3) covariate techniques. They correspond to methods applied during test construction, during test administration, and during data analysis, respectively.

Rational techniques prevent the respondent from answering in an unduly desirable fashion. For example, the test constructor can restrict item choice to those neutral in social desirability. Forced-choice items can be equated for social desirability. Demand reduction includes maximizing anonymity and confidentiality. If feedback is to be provided, respondents can be reminded that the feedback will be useful only if responses are honest. Of course, all such instructions must be provided *before* the test administration in order for them to reduce demand for socially desirable responses. Covariate techniques involve the administration of an SDR scale along with the content measure of interest. Scores on SDR are then partialed out of the content measure in an attempt to create a purified measure of content. We do not recommend the use of covariate techniques because they typically remove valid variance and, if anything, reduce the validity of the content measure (Ones, Viswesvaran, & Reiss, 1996). We do recommend that assessors use rational methods and demand reduction.

Note that concern over SDR may be unwarranted in many research contexts. In research on student or volunteer samples, there is little reason to be concerned about contamination from this response bias (Piedmont, McCrae, Riemannn, & Angleitner, 2000). SDR observed under such low-demand conditions is likely to reflect substantive variance, that is, traits with desirability implications.

STYLES AS TRAITS

Because consistent styles are likely to have their own cognitive or motivational roots, they can be studied as personality traits in their own right. As such, their manifestations are likely to go well beyond test-taking behaviors. The classic example is the Marlowe-Crowne scale. Ori-

ginally designed to measure socially desirable responding, the scale came to be interpreted in terms of need for approval (Crowne & Marlowe, 1964).

To further complicate the issue, repeated biases can eventually become honest self-reports: With enough repetition (Paulhus, 1993) or reinforcement (Jones, Davis, & Gergen, 1961), a self-presentation can be incorporated into the respondent's true self-image (Johnson & Hogan, 1981).

Acquiescent Responding (AR)

The label *acquiescent* is used for respondents who tend to agree with statements without regard to their content. Those with the opposite tendency, indiscriminant disagreement, are called *reactant*. Such tendencies are rarely absolute: Usually everyone agrees with some statements and disagrees with others. On dichotomous response formats (e.g., Yes–No, True–False, Agree–Disagree), the phenomenon is evident if respondents show dramatically high or low proportions of "Yes" answers across a wide range of items Another way of diagnosing either extreme is to index the tendency to agree with an affirmation and its exact negation (e.g., "happy" and "not happy").

The traditional concern is that acquiescence can be viewed as an individual difference variable in its own right—a personality trait with conceptual links to conformity and impulsiveness (see, e.g., Couch & Kenniston, 1960). If so, a problem arises when a self-report instrument measures acquiescence along with (or instead of) the construct it was designed to measure. For example, on most anxiety scales, the majority of items ask respondents to indicate which anxiety-related symptoms they have experienced. The respondent who agrees with all the symptoms may indeed be a very anxious person—or merely a yea-sayer. For this reason, some researchers have worried that acquiescence might be a serious confound in self-reports (Schuman & Presser, 1981). Other researchers have concluded that acquiescence effects are insignificant (Rorer, 1965).

In one sense, acquiescence is more problematic in attitude and survey research than in personality assessment (Krosnick, 1999). In survey research, the overall percentage agreement with an item (e.g., I favor capital punishment) is more critical than in personality items (e.g., I am friendly), where individual differences are 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 formats.

Instead, the major problem for personality research is that acquiescence exaggerates the correlations among same-valenced items and decreases the correlations among opposite-valenced items. Moreover, acquiescence in one personality domain correlates with acquiescence in other personality domains (Knowles & Nathan, 1997). Consequently, correlations between scales with items keyed in the same direction may be inflated and, conversely, two scales may appear orthogonal because their items are scored in opposite directions (McCrae, Herbst, & Costa, 2001; Paulhus, 1984).

A recent example of the substantive import of acquiescence is the debate about the bipolarity of affect. Carroll, Yik, Russell, and Barrett (1999) argued that AR artifactually reduces the correlation between positive and negative affect. A complex structural model was required to show that, in fact, the correlation remains modest even after taking into account the effects of acquiescence and therefore positive and negative affect can be measured separately (McCrae et al., 2001; Tellegen, Watson, & Clark, 1999).

CONTROL OF AR EFFECTS

The standard control for AR bias at the test construction stage is simply to balance the scoring key. That is, half the items should be written as true-keyed (a high rating indicates possession of the trait) and half the items are false-keyed (a low rating indicates possession of the trait).

This simple precaution controls the classical form of acquiescence (agreement acquiescence) because, to get an overall high content score, the respondent must agree with some items and disagree with others. Any effects of an acquiescent tendency on the true-keyed items will be canceled out on the false-keyed items. In other words, one cannot get a high content score simply by yea-saying or nay-saying (Wiggins, 1973).

An imbalanced key can even be dealt with post hoc. If the correlations are high between true- and false-keyed subtotals and their correlations with other variables are comparable, one may safely combine the two. If not, one could differentially weight the two subtotals to simulate a balanced key.

Two unfortunate problems arise from attempts to control acquiescence by balancing the key. One is a reduction in the alpha reliability of the instrument (as compared to one with unidirectional keying). Second, factor analyses often yield two factors—one for the true-keyed items and one for the false-keyed items—even when the underlying construct is unidimensional. The reason for both problems is that items keyed in the same direction tend to have higher correlations than items keyed in opposite directions.

MEASUREMENT OF AR

Only a handful of instruments have been designed to measure individual differences in AR (e.g., Couch & Kenniston, 1960) but none is widely used. A number of the larger assessment batteries permit computation of an acquiescence index across all the items in the battery. Gough's (1983) Adjective Check List (ACL), 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., Tracey, Rounds, & Gurtman, 1996). Note that this procedure may eliminate measurement of some content variables unless one has administered the ACL in True–False format to ensure some response to each item. The most credible measure of AR bias is an overall sum of items on a large relatively balanced personality inventory such as the NEO-PI-R (McCrae et al., 2001).

Extreme Responding (ER)

ER is the tendency to use the extreme choices on a rating scale (e.g., 1's and 7's on a 7-point scale). Situational factors such as ambiguity, emotional arousal, and rapid responding induce temporary increases in extreme responding. The individual exhibiting this tendency across time and stimuli may be said to have an extreme response style. Low scorers on this variable may be said to show moderate responding, that is, the tendency to use the midpoint as often as possible.

Early reviews (e.g., Peabody, 1962) concluded that ER bias is a consistent individual difference, and more recent studies have sustained this conclusion. ER bias appears to be highly stable over time and a major source of individual differences in raters. There is little support, however, for a link between ER bias and any traditional personality dimensions (Schwartz & Sudman, 1996).

Contamination of a dataset with ER bias precludes the direct comparability of one respondent'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 ER bias induces spurious correlations among otherwise unrelated constructs. A third source of problems is the interaction between ER bias and demographic variables such as gender, race, and education (Schuman & Presser, 1981).

CONTROL OF ER EFFECTS

ER bias cannot be corrected simply by balancing the key because extremity operates in both directions. Standardizing the within-subject variance equates subjects on extremity but subject variances are often inextricably confounded with subject means. In measuring self-esteem, for example, most responses are on the positive side of the rating scale, thus confounding high self-esteem and ER bias.

In some situations, ER bias can be controlled by rendering all items in a dichotomous format. After all, a "Yes" response is no more or no less extreme than a "No" response. The loss of reliability in moving from a multipoint to a dichotomous format can be compensated for by adding additional items.

Another approach is to require fixed distributions. For example, the respondent may be asked to give equal numbers of each response option. A more common forced distribution is a normal approximation: On a 5-point scale, for example, the required proportions of each rating would be 1(.05), 2(.10), 3(.70), 4(.10), 5(.05). Typing the item statements on small cards is often used to facilitate the adjustment of these proportions. This approach, called the Q-sort, capitalizes on the fact that it is much easier to judge the height of a stack of cards than it is to keep track of how many 5's one has used (Block, 1961).

MEASUREMENT OF ER BIAS

There are no standard instruments for assessing ER bias 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, Kroonenberg, & Bakker, 1985). Of course, this approach 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 ER bias from a measure of dimensional importance or salience for that topic.

Miscellaneous Response Sets

Other response biases—for example, pattern responding, random responding, and inconsistent responding—create less cause for concern. In pattern responding, participants simply mark their responses in a physical pattern (e.g., 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, etc., or all 3's). This phenomenon is best recognized by human eye, although some researchers have developed computer programs to recognize the most common of these. Random responding is more difficult to detect even by eye (Baer, Kroll, Rinaldo, & Ballenger, 1999). Both of these sets can be detected instead by including a subset of rare items in the subject's inventory (e.g., "I was born in Pago-Pago"; "I recently had a liver transplant"). Although all are possible, an accumulation of "Yes" responses to such items suggests that none of the respondent's answers can be trusted. These miscellaneous response sets are reviewed in more detail by Paulhus (1991).

The MMPI includes measures of several miscellaneous response biases (e.g., F, F-K, FBS). They are less common in batteries aimed at normal-range personality. The most well-researched measures of inconsistent and infrequent responding are available in the Multivariate Personality Questionnaire (Patrick, Curtin, & Tellegen, 2002; Tellegen, in press) and Personality Assessment Index (Morey, 1991).

Other Limitations

Constraints on Self-Knowledge

It is often assumed that an honest self-disclosure is sufficient to yield an accurate self-description that can outperform informant consensus and predict future behavior. According to this view, only response biases stand in the way of accuracy. The assumption is that there is only one "truth" about an individual, a truth that is fully available to that individual.

In fact, there are good reasons to believe otherwise. For one thing, much information is unavailable to the earnest self-assessor. Dunning, Heath, and Suls (2005) distinguish between information unavailable to the self-assessor and information that tends to be ignored by the self-assessor. People do not have an infinite ability to recall all information relevant to a posed question. Conversely, they may be overwhelmed with a plethora of information, in which case the process of integration and simplification may be too challenging a task. A self-reporter may have to resort to a "press release" version of his or her personality just to get on with the task.

Of more concern than lack of access is the claim that introspection may actually diminish accuracy. Timothy Wilson has argued that any extended attempt to clarify one's self-descriptions can undermine their validity. This effect may be restricted to gross evaluations of unfamiliar targets (e.g., Wilson & LaFleur, 1995).

Further research is needed to determine the relevance of this work to personality assessment. To date, we know of no such research on the effects of introspection on the validity of self-reports of personality. We do know that, up to a point, speeding the administration of a personality test has little effect on its validity (Holden, Wood, & Tomashewski, 2001).

Cultural Limitations

Respondents from different cultures may not treat self-reports in the fashion we expect from those of European heritage (Hamamura, Heine, & Paulhus, 2006). Those with Asian heritage, for example, show more moderacy bias and ambivalence (Chen, Lee, & Stevenson, 1995). Such stylistic differences may create artifactual differences between groups. When expected cultural differences are *not* found, the failure may sometimes be traced to the *reference group effect* (Heine, Lehman, Peng, & Greenholtz, 2002): That is, respondents normally evaluate themselves relative to their own culture and not to some unspecified external group. In principle, then, mean group differ-

ences should vanish or, at least, diminish in size.

Research on cultural differences in self-report styles has just begun and will become more complex as other cultural groups are studied in detail. In the meantime, we recommend that readers treat with caution any claims for cultural differences based purely on survey data.

Convergence with Other Methods

Because there is no absolute criterion against which a personality self-report can be evaluated, evidence for its construct validity must be marshaled from a variety of sources in a cumulative process called *construct validation* (Loevinger, 1957). Necessary for this process is support for convergent and discriminant validity. Convergent validity is advanced by the confirmation of associations among available self-report measures of the construct. If our new measure of extraversion correlates highly with the Extraversion scales on Eysenck's Personality Inventory and the Big Five Inventory, then potential users of the new instrument will have more faith that it is capturing the intended construct.

An even more impressive demonstration is the convergence among different modes of measurement. To the degree that a self-report of extraversion correlates highly with informant ratings, behavioral, and life data measures, then its credibility is boosted considerably. Established measures of the Big Five personality traits, for example, have demonstrated substantial convergence (in the .40–.60 range) with aggregated ratings of knowledgeable informants (e.g., McCrae & Weiss, Chapter 15, this volume). Research on moderators of self–other agreement can also give us some insight into the conditions under which self-reports are more likely to be accurate. For example, self–other agreement is high when the respondents are self-consistent and certain about the trait, and when the trait being rated is important to the respondent, unambiguous, more observable, and evaluatively neutral (John & Robins, 1993). Self–other agreement is higher for personality traits than for affective traits (Watson, Hubbard, & Wiese, 2000). Overall, the correspondence between self-reports and informant reports is moderate in size, suggesting that the two methods provide some overlapping and some complementary information (Vazire, 2006).

Convergence of self-reports with behavioral measures is more variable and depends on the degree of aggregation, the reliability, and the relevance of the behavioral measures. Research has found that self-behavior convergence is higher for affect-related traits (Spain, Eaton, & Funder, 2000) and for evaluatively neutral behaviors (Gosling, John, Craik, & Robins, 1998). Overall, the relation between self-reports and behavior tends to be modest (Meyer et al., 2001; Vazire, 2006). However, as Meyer and colleagues (2001) point out, those correlations are within the range of other well-established real-world effects (e.g., mammogram results predicting breast cancer).

Convergence with other modes of measurement can reduce concerns about *common method variance*. The term refers to possible contamination ensuing from the use of a single measurement method: It can exaggerate the apparent association between two constructs measured with the same method (Wiggins, 1973). These concerns are often justified because large datasets are often composed entirely of self-report measures. Response biases common to self-reports (see section above) may distort the intercorrelations. Associations among self-report measures of the Big Five, for example, may be exaggerated by the operation of self-favoring biases (McCrae & Costa, 1989) and acquiescence (McCrae et al., 2001). Spurious associations may result when self-reports are used to measure predictors such as coping, defensiveness, or self-enhancement as well as adjustment outcomes (Colvin, Block, & Funder, 1994; Cramer, 1998).

As noted earlier, in some assessment situations, the self-report mode of measurement is the most credible and is therefore used as the criterion method. The literature on so-called *zero acquaintance*, for example, examines the increases in the validity of observer ratings with increases in level of acquaintanceship. The size of the association with a corresponding self-report measure is used to index the validity of an observer judgment (Kenny, 1994).

Selecting a Self-Report Instrument

Having selected the construct to be measured and having concluded that self-report is the

method of choice, the researcher still has a series of decisions to make.

Established or New Instrument?

As a general rule, the researcher should use an established instrument rather than one with less scientific credibility. An established measure is likely to have undergone an extensive program of construct validation. Normative data are also more likely to be available. A comparison of one's results with norms helps ensure that one hasn't miscored or misapplied the instrument.

If no credible instrument is available, one might have to construct a new one. This task requires a lengthy and often expensive program of construct validation (see Simms & Watson, Chapter 14, this volume). Without such validation, the use of an ad hoc measure is vulnerable to criticism. Any association (or nonassociation) found with the measure can be explained away as a faulty operationalization of the construct.

Which One?

The choice of an established instrument will require substantial homework and consultation with experts. If one seeks to cover the broad domain of personality, a number of well-researched inventories are available. Many of them have organized personality into the well-known Five-Factor Model alternatively known as the "Big Five."

Only a few instruments provide a multifaceted breakdown of the Big Five factors. One is the commercially available NEO Personality Inventory—Revised (NEO-PI-R; Costa & McCrae, 1989), the most widely used. Another is the International Personality Item Pool (IPIP) inventory, which can be freely downloaded from Lewis Goldberg's website (*www.ipip.ori.org/*; Goldberg et al., 2006). A sixth factor (Honesty-Humility) is included along with the Big Five factors in the HEXACO measure; facet scales are included for all six factors (Lee & Ashton, 2004). The Hogan Personality Inventory (HPI; Hogan & Hogan, 1992) comprises seven factors and was the first to subdivide major personality dimensions into meaningful facets. Finally, the Five Factor Inventory (Hofstee & de Raad, 2002) was developed in Dutch, but with easy translation into other languages in mind.

Several other instruments are much shorter, with the more modest goal of capturing the Big Five factors without the facets. These include John and Srivastava's (1999) Big Five Inventory (BFI), Costa and McCrae's (1989) NEO Five-Factor Inventory (NEO-FFI), and Goldberg's (1990) Trait Descriptive Adjective (TDA) markers. Saucier (1994) has developed an abbreviated set of adjective minimarkers.

Other comprehensive instruments organize the personality space using a variety of alternative schemes. These include Gough's (1957/1995) California Personality Inventory (CPI), Block's (1961) Q-Set, Cattell's (Cattell & Schuerger, 2003) 16PF, Jackson's (1984) Personality Research Form (PRF) and Tellegen's (in press) Multidimensional Personality Questionnaire (MPQ) (see Patrick, Curtin, & Tellegen, 2002).

Despite the availability of all the variables measured in those inventories, new constructs are proposed and measured on a regular basis. Instead of writing original items, researchers sometimes start with one of the comprehensive item sets cited above. These inventories contain a wide enough variety of items for use in developing new personality measures. For example, a set of experts might be asked to rate the 100 Q-Set items for prototypicality with regard to sanctimoniousness. The items with the highest mean ratings could then be assembled to form a new Sanctimone Scale. An advantage to this approach is that the new measure can be scored on archived datasets that include the full inventory (see Cramer, Chapter 7, this volume).[3]

The Eysenck Personality Questionnaire (EPQ) may well have been administered more often than any other inventory of normal personality, but only three variables can be scored (Extraversion, Neuroticism, Psychoticism). Several other broad instruments are organized in terms of temperament—for example, the Tridimensional Character Inventory (Cloninger, 1987) and the EAS (Buss & Plomin, 1984). Other narrower domain instruments include Block's (1965) ego-control and ego-resiliency scales. Several focus specifically on maladaptive personality traits: the Dark Triad (Paulhus & Williams, 2002) and the Hogan Personality Factors measure (Hogan & Hogan, 2001).

Single Variables

Although researchers are often interested in studying the role of a specific personality vari-

able, we recommend that it be measured along with a careful selection of other variables—especially those that may provide an alternative interpretation of the findings. Inclusion of parallel measures addresses convergent validity, and competing measures can provide discriminant validity. Critics want to know the incremental validity of the focal instrument. What does it capture above and beyond well-established traits? For example, a researcher claiming to study the role of self-esteem or coping must show that self-reports of those variables cannot be fully explained by other traits (e.g., neuroticism). Common these days is the inclusion of an instrument capturing all of the Big Five factors. All of these procedures reduce the possibility of researchers "reinventing the wheel."

In short, single-variable research is not recommended in isolation. The obvious tradeoff is with the space and time it takes to administer corroborative measures. Without such corroboration, however, interpretation of results with the key variable may remain ambiguous.

There are literally hundreds of single-construct self-report personality measures. Less than 50, we estimate, have sufficiently documented construct validity to be taken seriously. Two popular collections of established personality tests are the handbook by Robinson, Shaver, and Wrightsman (1991) and the online Social-Personality Psychology Questionnaire Instrument Compendium compiled by Reifman (2006). Reviews of commercially available instruments are provided in the venerable series of *Mental Measurement Yearbooks* (Plake, Impara, & Spies, 2003).

Summary

There are several indisputable advantages to the self-report method. It opens a pipeline to prodigious amounts of unique information about the target of assessment. It directly taps his or her self-perceived personality, that is, identity. Clarity of communication and ease of administration are also clear advantages. More than other methods, self-reports allow for the collection of large numbers of personality-relevant variables in one administration.

The disadvantages of self-reports have been given much scrutiny, especially in regard to response styles such as socially desirable responding. It is only prudent to be skeptical about respondents' claims about their personality—especially on highly evaluative traits. Although self-reports of personality can be faked, it is rarely a serious problem in most research settings. In high-stakes testing such as job interviews, self-presentation remains an issue. As with the use of any method, self-reports should be corroborated with alternative assessment methods.

Nonspecialist researchers planning to use self-reports can profit from choosing a well-established instrument: The necessary but arduous accumulation of psychometric credibility has already been carried out for measures of many constructs. The use of such measures will allow researchers to build on a cumulative science.

Well-constructed self-report scales can predict a wide range of important outcomes with ease and efficiency. Although relentlessly criticized, they remain the most popular means of personality assessment. We conclude by borrowing Winston Churchill's comparison of democracy to other political systems: As a method for accurate personality assessment, self-report is dreadful—yet, overall, more effective than any alternative.

Notes

1. In fact, the alpha reliability cannot even be calculated with a single item. It must be estimated from previous research in which that item was included with similar others.
2. Funder (2004) describes this approach as collecting behavioral data via self-report data.
3. Note that inventories sold commercially usually place legal restrictions on what items can be used in new measures. Nonetheless, one learn from them what type of item taps the new construct and then devise similar but noncopyrighted items.

Recommended Readings

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing oneself.* New York: Psychology Press.

Lucas, R. E., & Baird, B. M. (2006) Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29–42). Washington, DC: American Psychological Association.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and so-*

cial psychological attitudes (pp. 17–59). New York: Academic Press.

Schwarz, N., & Sudman, S. (Eds.). (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco: Jossey-Bass/Pfeiffer.

Tourangeau R., Rips, L. J., & Rasinski, K. (2001). *The psychology of survey response.* New York: Cambridge University Press.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press/Harvard University Press.

References

Adams, D. P. (2000). *The person: An integrated introduction to personality psychology.* Fort Worth, TX: Harcourt.

Allport, F. H. (1927). Self-evaluation: a problem in personal development. *Mental Hygiene, 11,* 570–583.

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308–320.

Block, J. (1961). *The Q-Sort method in personality assessment and psychiatric research.* Springfield, IL: Charles C. Thomas.

Block, J. (1965). *The challenge of response sets.* New York: Century.

Burisch, M. (1983). Approaches to personality inventory construction: A comparison of merits. *American Psychologist, 39,* 214–227.

Buss, A. H., & Plomin, R. (1984). *Temperament.* Boston: Allyn-Bacon.

Butcher, J. N. (2003). Computerized psychological assessment. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 141–163). Hoboken, NJ: Wiley.

Carroll, J. M., Yik, M. S. M., Russell, J. A., & Barrett, L. F. (1999). On the psychometric principles of affect. *Review of General Psychology, 3,* 14–22.

Cattell, H. E. P., & Schuerger, J. M. (2003). *Essentials of 16PF assessment.* New York: Wiley.

Cheek, J. M., & Watson, A. K. (1989). The definition of shyness: Psychological imperialism or construct validity? *Journal of Social Behavior and Personality, 4,* 85–95.

Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among east Asian and North American students. *Psychological Science, 6,* 170–175.

Cloninger C. R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry, 44,* 573–580.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68,* 1152–1162.

Costa, P. T., & McCrae, R. R. (1989). *Manual for the NEO Personality Inventory: Five Factor Inventory/ NEO-FFI.* Odessa, FL: PAR.

Couch, A., & Kenniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology, 60,* 151–174.

Cramer, P. (1998). Coping and defense mechanisms: What's the difference? *Journal of Personality, 66,* 919–931.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentational dynamics on overt integrity tests: Experimental studies of the Reid Report. *Journal of Applied Psychology, 79,* 643–658.

Davidson, K., & MacGregor, M. W. (1998). A critical appraisal of self-report defense mechanism measures. *Journal of Personality, 66,* 965–992.

Diener, E., Sandvik, E., Pavot, W., & Gallagher, D. (1991). Response artifacts in the measurement of subjective well-being. *Social Indicators Research, 24,* 35–56.

Dunning, D., Heath, C., & Suls, J. M. (2005). Flawed self-assessment: Implications for education, and the workplace. *Psychological Science in the Public Interest, 5,* 69–106.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality, 26,* 103–120.

Funder, D. C. (2004). *The personality puzzle.* San Francisco: Norton.

Goldberg, L. R. (1990). The structure of phenotypic personality traits. *American Psychologist, 48,* 26–34.

Goldberg, L. R. (1992). The development of markers for the Big Five structure. *Psychological Assessment, 4,* 26–42.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., et al. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74,* 1337–1349.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality, 37,* 504–528.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies?: A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

Gough, H. G. (1983). *The adjective check list.* Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1995). *Manual for the California Psychological Inventory* (3rd ed.). Palo Alto, CA: Consulting Psychologists Press. (Original work published 1957)

Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voter behavior by asking people if they expect to vote. *Journal of Personality and Social Psychology, 72*, 315–318.

Hamamura, T., Heine, S. J., & Paulhus, D. L. (2006). *Cultural differences in response styles: The role of dialectical thinking.* Manuscript under review.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group problem. *Journal of Personality and Social Psychology, 82*, 903–918.

Hofstee, W. K. B., & de Raad, B. (2002). The Five-Factor Personality Inventory: Assessing the Big Five by means of brief and concrete statements. In B. de Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 79–100). Amsterdam: Hogrefe & Huber.

Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory manual.* Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment, 9*, 40–51.

Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621–626.

Hogan, R., & Smither, R. (2001). *Personality: Theory and applications.* Boulder, CO: Westview Press.

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.

Holden, R. R., Wood, L. L., & Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology, 81*, 160–169.

Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. A. Johnson & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 165–195). San Diego, CA: Academic Press.

Jackson, D. N. (1984). *Personality Research Form manual.* London, ON, Canada: Research Psychologists Press.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521–551.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality psychology* (pp. 102–139). New York: Guilford Press.

Johnson, J. A. (2004). Impact of item characteristics on item and scale validity. *Multivariate Behavioral Research, 39*, 273–302.

Johnson, J. A., & Hogan, R. (1981). Moral judgments and self-presentations. *Journal of Research in Personality, 15*, 57–63.

Jones, E. E., Davis, K. E., & Gergen, K. J. (1961). Some determinants of being approved or disapproved as a person. *Journal of Abnormal and Social Psychology, 63*, 302–310.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford Press.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*, 311–337.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual process theory of acquiescence. *Journal of Personality and Social Psychology, 77*, 379–386.

Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern? *Journal of Research in Personality, 31*, 293–301.

Krosnick, J. A. (1999). Maximizing questionnaire quality. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitude* (pp. 37–57). San Diego, CA: Academic Press.

Kuhn, M. H., & McPartland, T. S. (1954). An empirical investigation of self-attitudes. *American Sociological Review, 19*, 68–76.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research, 39*, 329–358.

Little, B. R. (1998). Personal project pursuit: Dimensions and dynamics of personal meaning. In P. T. Wong & P. Fry (Eds.), *The human quest for meaning: A handbook of psychological research and clinical applications* (pp. 193–212). Hillsdale, NJ: Erlbaum.

Lucas, R. E., & Baird, B. M. (2006). Global self-assessment. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29–42). Washington, DC: American Psychological Association.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(Suppl. 9), 635–694.

McAdams, D. P. (2000). *The person. An integrated introduction to personality psychology.* Fort Worth, TX: Harcourt.

McCrae, R. R., & Costa, P. T. (1989). Different points of view: Self-reports and ratings in the assessment of personality. In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology: An interactional perspective* (pp. 429–439). Amsterdam: Elsevier.

McCrae, R. R., Herbst, J. H., & Costa, P. T. (2001). Effects of acquiescence on personality factor structures. In R. Reimann, F. M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 216–231). Berlin: Pabst Science.

Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1998). Socially desirable responding

and sexuality self-reports. *Journal of Sex Research, 35*, 148–157.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*, 128–165.

Mills, C., & Hogan, R. (1978). A role theoretical interpretation of personality scale item responses. *Journal of Personality, 46*, 778–785.

Morey, L. C. (1991). *The Personality Assessment Inventory Professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C., & Lanier, V. W. (1998). Operating characteristics of six response distortion indicators for the Personality Assessment Inventory. *Assessment, 5*, 203–214.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.

Osberg, T. M. (1999). Comparative validity of the MMPI-2 Wiener–Harmon subtle–obvious scales in male prison inmates. *Journal of Personality Assessment, 72*, 36–48.

Ozer, D. J., & Reise, S. P. (1994). Personality assessment. *Annual Review of Psychology, 45*, 357–393.

Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment, 14*, 150–163.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic Press.

Paulhus, D. L. (1993). Bypassing the will: The automatization of affirmations. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 373–587). Englewood Cliffs, NJ: Prentice-Hall.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67–88). Hillsdale, NJ: Erlbaum.

Paulhus, D. L., & & Levitt, K. (1986). Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology, 52*, 245–259.

Paulhus, D. L., & Williams, K. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*, 341–350.

Peabody, D. (1962). Two components in bipolar scales: Direction and extremeness. *Psychological Review, 69*, 65–73.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC 2001).* Mahwah, NJ: Erlbaum.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582–593.

Plake, B. S., Impara, M. C., & Spies, R. A. (Eds.). (2003). *The Fifteenth Mental Measurements Yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Raskin, R. N., & Hall, C. S. (1981). The narcissistic personality inventory: Alternative form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*, 159–162.

Reifman, A. (2006). *Social-Personality Psychology Questionnaire Instrument Compendium.* El Paso: University of Texas at El Paso. Available at *www.hs.ttu.edu/research/reifman/qic.htm.*

Roberts, B. W., O'Donnell, M., & Robins, R. W. (2004). Goal and personality trait development in emerging adulthood. *Journal of Personality and Social Psychology, 87*, 541–550.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*, 151–161.

Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649–679). San Diego, CA: Academic Press.

Robins, R. W., Norem, J. K., & Cheek, J. M. (1999). Naturalizing the self. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 443–477). New York: Guilford Press.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). *Measures of personality and social psychological attitudes.* San Diego, CA: Academic Press.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677–688.

Rorer, L. G. (1965). The great response-style myth. *Psychological Bulletin, 63*, 129–156.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63*, 506–516.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys.* New York: Academic Press.

Schwartz, N., & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research.* San Francisco: Jossey-Bass.

Sedikides, C., & Strube, M. J. (1995). The multiply motivated self. *Personality and Social Psychology Bulletin, 21*, 1330–1335.

Spain, J. S., & Eaton, L. G., & Funder, D. C. (2000).

Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality, 68*, 837–867.

Swann, W. B., Jr., Chang-Schneider, C., & McClarty, K. L. (in press). Do our self-views matter?: Self-concept and self-esteem in everyday life. *American Psychologist.*

Tellegen, A. (in press). *Multidimensional Personality Questionnaire.* Minneapolis: University of Minnesota Press.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10*, 297–303.

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775–801.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2001). *The psychology of survey responses.* New York: Cambridge University Press.

Tracey, T. J., Rounds, J., & Gurtman, M. (1996). Examination of the general factor with the interpersonal circumplex structure: Application to the Inventory of Interpersonal Problems. *Multivariate Behavioral Research, 31*, 441–466.

Trapnell, P. D. (2006). *A new measure of agentic and communal values.* Manuscript under review.

Trapnell, P. D., & Campbell, J. D. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology, 76*, 284–304.

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.

Vazire, S. (2006). Informant reports: A cheap, fast, easy method for personality assessment. *Journal of Research in Personality, 40*, 472–481.

Vazire, S., & Gosling, S. D. (2004). e-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123–132.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*, 546–558.

Wiggins J. S. (1973). *Personality and prediction: Principles of personality assessment.* Boston: Addison-Wesley.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, MA: Belknap Press/Harvard University Press.

Wilson, T. D., & LaFleur, S. J. (1995). Knowing what you'll do: Effects of analyzing reasons on self-prediction. *Journal of Personality and Social Psychology, 68*, 21–35.

Woods, S. A., & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality, 19*, 373–390.