HUMAN PERFORMANCE, *11*(2/3), 245–269

The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection

Deniz S. Ones
Department of Psychology
University of Minnesota

Chockalingam Viswesvaran
Department of Psychology
Florida International University

A review of the extant literature and new empirical research suggests that social desirability is not much of a concern in personality and integrity testing for personnel selection. In particular, based on meta-analytically derived evidence, it appears that social desirability influences do not destroy the convergent and discriminant validity of the Big Five dimensions of personality (Emotional Stability, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness). We also present new empirical evidence regarding gender and age differences in socially desirable responding. Although social desirability predicts a number of important work variables such as job satisfaction, organizational commitment, and supervisor ratings of training success, social desirability does not seem to be a predictor of overall job performance and is only very weakly related to specific dimensions of job performance such as technical proficiency ($r = -.07$) and personal discipline ($r = .05$). Large sample investigations of the moderating influences of social desirability in actual work settings indicate that social desirability does not moderate the criterion-related validities of personality variables or integrity tests. The criterion-related validity of integrity tests for overall job performance with applicant samples in predictive studies is .41. Controlling for social desirability in integrity or personality test scores leaves the operational validities intact, thereby suggesting that social desirability functions neither as a mediator nor as a suppressor variable in personality–performance and

Requests for reprints should be sent to Deniz S. Ones, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455–0344. E-mail: Deniz.S.Ones-1@tc.umn.edu

integrity–performance relations. Theoretical explanations of why social desirability does not influence criterion-related validity are reviewed.

Since the early 1990s personality measurement and integrity testing for industrial and organizational applications has been on the rise (Hough & Schneider, 1996). The increasing credibility of personality measurement in Industrial/Organizational (I/O) psychology is the result of large-scale studies and meta-analyses reporting substantial validities for theoretically relevant personality constructs for various criteria, including supervisory ratings of job performance, contextual performance, counterproductive behaviors at work, drug and alcohol abuse, and violence at work (e.g., Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy 1990; Ones, Viswesvaran, & Schmidt, 1993; Tett, Jackson, & Rothstein, 1991). Despite increasing acceptability of personality measures among human resources managers, I/O psychologist practitioners, and researchers, there are still a number of concerns voiced about the use of personality measures in personnel selection. In particular, potential social desirability among job applicants is often cited as a concern about using personality scales and integrity tests in applied settings. In this article, we review existing evidence and present new evidence that social desirability does not influence the validity of personality measures or integrity tests.

Our purpose is to provide an overview of the impact of social desirability on psychometric properties and practical uses of personality scales and integrity tests. Response sets have been and continue to be the most frequently cited criticism of personality and integrity testing for personnel selection (Block, 1965; Murphy & Davidschofer, 1998). Sometimes, practitioners are hesitant to use integrity tests for fear that the scores may be altered or improved by faking. In this article, we first review the impact of social desirability on personality scale scores. Second, we examine the impact of social desirability on convergent and divergent validity of personality measures. Third, we examine gender and age differences on social desirability scales. Fourth, we direct our attention to social desirability and criterion-related validity. This article is intended to review existing evidence on the impact of social desirability in personnel selection and to present new empirical research to enhance our understanding of personality measurement, integrity testing, and faking.

IMPACT OF FAKING ON SCALE SCORES

How does social desirability influence basic psychometric properties of personality scales and integrity tests? This question has been under scrutiny since the 1930s (Kelly, Miles, & Terman, 1936; Hough, in press). The most frequently used method in examining differences in distributional properties of personality scores when individuals distort their responses has been the use of faking instructions to induce socially desirable responding and to compare the score distributions with those

obtained under honest response conditions. Studies examining the susceptibility of personality inventories to faking have employed either a within-subjects or a between-subjects experimental design (Furnham, 1986). In the within-subjects experimental designs, the same individuals take the personality inventory under two instructional sets. The responses of the same individuals across the two instructional sets are compared. In the between-subjects experimental design, the responses of one group of individuals instructed to fake (either good or bad) is compared to the responses obtained from another group of individuals instructed to answer honestly. The two designs have their advantages and disadvantages (Cook & Campbell, 1979). In terms of statistical power, given equal number of subjects, the within-subjects design is more powerful. More important, the validity of the between-subject designs is predicated on the equivalence of the two groups and that there is no instruction-by-subject interaction.

In any event, a comparison of scale score distributions under faking instructions and honest response conditions addresses the question of whether or not individuals can fake their responses on personality scales, if instructed to do so. Comparison of mean scale scores for the different faking instructional sets examines the maximal limits on fakability of various personality scales. It is important to note, however, that a finding that individuals can fake does not necessarily imply that they do in real-world applications (Hough & Schneider, 1996; McCrae & Costa, 1983; Schwab, 1971). Although very useful from the point of examining social desirability dynamics, faking studies are silent about actual faking in real-world situations.

Impact of Faking on Personality Scale Scores

Viswesvaran and Ones (in press) used meta-analytic cumulation to answer the question of whether or not individuals can fake their responses on personality inventories if instructed to do so. The Viswesvaran and Ones study examined mean scale score differences under fake good instructions. The results indicated that if instructed to fake good, the respondents were able to change their responses by almost .50 standard deviations on the Big Five factors. Within-subjects design produced larger effect sizes and greater variability across the Big Five factors than the between-subjects designs. Across the Big Five personality dimensions, the standardized mean differences between individuals instructed to fake good and respond honestly were .72 for within-subjects studies and .60 for between-subjects design studies. In other words, participants can increase their scores by over .50 standard deviations on personality scales, if instructed to do so.

Impact of Faking on Integrity Test Scores

There are over 15 similarly conducted individual faking studies for integrity tests (e.g., Cunningham, Wong, & Barbee, 1994; Dean, 1990; Lobello & Sims, 1990;

Moore, 1990; Ryan & Sackett, 1987; Wanek, 1991). Similar to the findings for personality scales, the results from these studies indicate that individuals instructed to represent themselves in a favorable light can do so. On average, a comparison of individuals instructed to fake good on integrity tests compared with those instructed to respond honestly indicate effect sizes over .50 (observed mean = .86, observed standard deviation = .8132; total N = 1,626). However, it is again crucial to realize that faking integrity tests under laboratory instructions does not indicate the level of actual faking in real-world job applicant situations (to the extent that it exists). Further, increases in mean test scores that might be produced by situational demands may not be relevant to construct and criterion-related validity.

Impact of Faking on Social Desirability Scale Scores: Do Social Desirability Scales Capture Faking?

An interesting finding from the Viswesvaran and Ones (in press) study of faking effect sizes was the large effect sizes found for social desirability scales. That is, the largest mean differences between "fake good" and "respond honestly" instructional sets were found on the social desirability scales. For between-subjects design studies, the observed d value was 1.06 when fake good condition responses were compared to responses under honest response instructions. The corresponding value for within-subjects design studies was 2.26 standard deviation units. When the responses from participants instructed to fake bad were compared to responses from participants instructed to respond honestly, the mean observed d value for between-subjects design studies was found to be 1.17, while it was 3.66 for within-subjects design studies. Table 1 summarizes Viswesvaran and Ones's (in press) results for fakability estimates of social desirability scales.

Recall that Viswesvaran and Ones's (in press) meta-analysis indicated the mean fakability estimates (d values) across substantive personality scales were .72 for

TABLE 1
Influences of Faking (Good and Bad) Instructions on Social Desirability Scale Scores

Comparison	*Study Design*	*K*	*N*	*Mean d*
Fake good–honest	Between-subjects	26	2,023	1.06
Fake good–honest	Within-subjects	19	609	2.26
Fake bad–honest	Between-subjects	34	1,712	–1.17
Fake bad–honest	Within-subjects	29	751	–3.66

Note. Selectively summarized from "Meta-analysis of fakability estimates: Implications for personality measurement," by C. Viswesvaran and D. S. Ones, in press. Copyright 1998 by Sage. Adapted with permission. K = Number of d values being pooled; Mean d = standardized difference between the means of the two groups identified in the first column.

within-subjects studies and .60 for between-subjects design studies. It appears that fake good instructions produce larger increases in response distortion scale scores rather than in substantive personality scale scores. Across the board, scores changed more dramatically in social desirability scale scores than in any other content-oriented personality scale.

There are three important implications from the aforementioned finding. The first implication is that response distortion scales are likely to be useful in flagging individuals who fake, both in the socially desirable and in the socially undesirable direction. The large effect sizes indexing the mean differences between positively distorted and honest social desirability scale scores point to minimal overlap between the distributions. More significantly, coupled with Viswesvaran and Ones's (in press) finding of substantially smaller fakability estimates for substantive personality scales, these results suggest that practitioners can use response distortion scales to identify individuals who may be distorting responses in personnel selection situations. Social desirability scales appear to be very sensitive to response distortion.

Second, social desirability scales are likely to be useful in capturing faking. Based on within-subjects fakability estimate for social desirability scales (*d* value for fake good vs. respond honestly instructional sets), our best estimate of the observed mean correlation between a dichotomous criterion of faking and a social desirability scale is .75. Because faking is a continuous variable, in that even under instructions to fake not all study participants fake equally, correcting the faking criterion for dichotomization (Hunter & Schmidt, 1990a) results in a correlation of .95 between a continuous faking criterion and social desirability scale scores. In light of this high observed correlation based on multiple studies, it is hard not to conclude that social desirability scales capture faking very well.

The third implication of usefulness of social desirability scales in identifying fakers concerns simulation studies in this domain. Simply put: Simulation studies should focus on realistic ranges of true score distortion. A simulation study that concludes faking is a problem when the true scores range from –4.0 to +4.0 standard deviations and that faking is assumed to change true scores by 5 standard deviations is not meaningful. In fact, a simulation study that examines whether rank ordering is affected when faking increases scores by 1 standard deviation does not simulate the reality in fake-good studies, let alone the reality in personnel selection, and should probably not be used in making decisions about the use of personality scales in personnel selection. In fact, we have reservations about the use of simulations to investigate the effects of faking in general. There are two main problems here. First, the conclusions from simulation studies, although couched in terms of personality measurement, are equally applicable to interviews and other noncognitive predictors. Second, it is probably very naive to use linear transformations of true scores to model faking in real-world situations. As we point out in the next section of this article, this is because social desirability is related to real differences in personality

scale scores. We next turn our attention to social desirability influences on construct validity of personality variables.

SOCIAL DESIRABILITY AND CONSTRUCT VALIDITY

Examinations of construct validity can involve investigations of convergent and divergent validity as well as factor structure. We take up each of these essential and complimentary ways of studying construct validity in turn.

Convergent and Divergent Validity

The influences of social desirability on convergent and divergent validity of personality scales has not been as extensively studied. A demonstration of convergent and divergent validity requires that scales designed to measure the same construct correlate more highly among themselves than with scales designed to measure other constructs (Campbell & Fiske, 1959). Based on Ones (1993) and Ones, Schmidt, and Viswesvaran, (1994), Ones, Viswesvaran, and Reiss (1996) reported the meta-analytically obtained matrix of intercorrelations among the Big Five dimensions of personality; 5,703 correlations contributed to the analyses (total N = 4,193,974). The relations reported were corrected for attenuation in both variables being correlated. These meta-analytic results indicated that correlations between corresponding Big Five dimension scales are higher than those between scales of different dimensions. In other words, the highest correlations in the table of Big Five intercorrelations are in the within-category diagonal. Scales from quite different personality inventories tapping the same dimension of the Big Five correlated more highly than those tapping into other dimensions of the Big Five. What is the impact of social desirability on the convergent and divergent validity of the Big Five?

To answer this question, social desirability was partialled out from the correlations among the Big Five personality dimensions. The intercorrelations between social desirability and the Big Five were taken from Ones et al. (1996). The results are reported in Table 2.

The results indicate that the convergent validities for each of the Big Five dimensions do not change by much when social desirability is partialled out. The largest decrease in convergent validity was found for emotional stability. For this personality dimension, convergent validity prior to partialling social desirability was .63. When social desirability is partialled out, convergent validity for emotional stability is .57. For conscientiousness, the decrease in social desirability was .02, a drop from .47 to .45. Note that discriminant validities in Table 2 are mostly unaffected by social desirability influences (an average decrease of .015 in correlations). Further, the pattern of partial correlations confirms that data from many diverse personality inventories fit the Big Five factor structure of personality. These

TABLE 2
Social Desirability Influences on Convergent and Divergent Validity of the Big Five Personality Dimensions

Personality Dimension	*1 – r*	*1 – Partial r*	*2 – r*	*2 – Partial r*	*3 –r*	*3 – Partial r*	*4 – r*	*4 – Partial r*	*5 – r*	*5 – Partial r*
Emotional Stability	.63	.57								
Extraversion	.19	.18	.41	.41						
Openness to Experience	.16	.17	.17	.17	.43	.43				
Agreeableness	.25	.22	.17	.16	.11	.11	.53	.52		
Conscientiousness	.26	.20	.00	–.01	–.06	–.06	.27	.25	.47	.45

Note. The intercorrelations among the Big Five and covergent validities for each of the Big Five from "The role of social desirability testing for personnel selection: The red herring," by D. S. Ones, C. Viswesvaran, and A. D. Reiss, 1996, *Journal of Applied Psychology, 81,* pp. 660–679.

meta-analytically based results lead us to conclude that convergent and divergent validities of the Big Five dimensions of personality are virtually unaffected by social desirability influences.

Our results dovetail the convergent and divergent validities of personality scales used with real job applicants. Convergent and divergent validities of personality inventories from both Big Five-based (e.g., Hogan Personality Inventory [HPI]) and non Big Five-based (e.g., California Psychological Inventory [CPI]) frameworks for job applicant samples have been reported. For example, the HPI manual (Hogan & Hogan, 1995) reports correlations among HPI scales and a number of other personality measures, including Inventory of Personal Motives (Hogan & Jones, 1992), Interpersonal Adjective Scales (Wiggins, 1991), Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), and PROFILE (Jones, 1990). Data from a total of 5,506 job applicants were used in computing convergent and divergent validities, which indicate that any social desirability influences that may have acted among job applicants *do not destroy* these two important indices of construct validity.

Factor Structure

In examining construct validity, it is common to examine the factor structure, as well as convergent and discriminant validity. There have been a number investigations of social desirability influences on the factor structure of personality inventories in applied settings (Cellar, Miller, Doverspike, & Klawsky, 1996; Hogan & Hogan, 1991; Livneh & Livneh, 1989) as well as direct factor structure comparisons between applicant and nonapplicant groups (Michaelis & Eysenck, 1971; Schmit & Ryan, 1993.) Unfortunately, the findings from this stream of this research are not unequivocal.

For example, Cellar et al.'s (1996) confirmatory factor analyses of two Big Five measures (Goldberg Five Factor Markers, Goldberg, 1992; and NEO–Personality Inventory, Costa & McCrae, 1985) using data from flight attendant trainees indicated a better fit for a six-factor solution to the data, compared to a five-factor solution. Hogan and Hogan's (1991) principal components analysis of another Big Five-based inventory (the HPI), based on data from employees, revealed the presence of eight components. Livneh and Livneh (1989) used a non Big Five inventory, the Adjective CheckList (ACL; Gough & Heilbrun, 1983), in gathering data from human service providers. They were unable to extract a five-factor solution. Michealis and Eysenck (1971) compared the factor structures and factor intercorrelations for personality scale scores using data from both job applicants and nonapplicants. Differences between job applicants and nonapplicants were found in factor pattern matrices and factor intercorrelations. More recently, Schmit and Ryan (1993) also compared the factor structures and factor intercorrelations for a Big Five inventory (NEO–PI) for both job applicants and nonapplicants

(students). While the five-factor solution fit the data from nonapplicants, exploratory factor analysis indicated a better fit for a six-factor solution to the applicant data. This sixth factor was labeled *an ideal employee* factor. Also, the factor intercorrelations in the applicant sample were larger than those obtained for nonapplicants. Finally, Costa (1996) interpreted the results from these investigations as follows: "Effects of evaluation bias on the structure of the NEO–FFI are relatively modest" (p. 231). While we tend to agree, we also think what is needed is the reporting or availability of full intercorrelation or covariance matrices on which factor analyses are based. Only using multiple applicant and multiple nonapplicant samples, we can rule out sample specific influences and sampling error as the source of the differences that have been reported across the factor structure examinations. In this regard, the methodology of basing factor analyses on meta-analytically derived intercorrelation matrices may help (Viswesvaran & Ones, 1995).

GENDER AND AGE DIFFERENCES ON SOCIAL DESIRABILITY: THE RESULTS OF A META-ANALYSIS

It has been hypothesized that there are gender differences in faking personality inventories (e.g., Hammill & Wheeler, 1997; Gannon, Raber, Jenkins, Ketterman, & Griffith, 1997). Critics have argued that because of gender differences in social desirability, women might "have less of a chance to be hired" (Gannon et al., 1997).

In this portion of our article, our aim is to investigate gender and age differences in social desirability. Study results examining gender differences in social desirability scale scores and study results reporting social desirability–age correlations were quantitatively cumulated using the methods of psychometric meta-analysis. Such an investigation is important because a number of personality inventories, most notably the 16PF, recommend the use of score adjustments to scale scores based on responses to social desirability scales. The Civil Rights Act of 1991 prohibits different cutoffs for minorities and women. To the extent that there are mean score differences between men and women, adjusting substantive personality scale scores based on social desirability scores could lead to the creation of adverse impact for protected groups.

We searched the literature for studies reporting means and standard deviations on social desirability scales for male and female samples separately. We also searched for correlations, or information that could allow the computation of a correlation, between social desirability measures and age. The PsycLIT database was searched for studies addressing social desirability for the time period 1974 to 1995. A manual search of the psychological abstracts was also undertaken, covering the years 1945 to 1973. A snowballing technique, whereby the references of the obtained articles were searched to identify further relevant studies, was also used. The social desirability scales represented in our database came from the

following personality inventories: Assessment of Background and Life Experiences (ABLE), CPI, 16PF, ACL, MMPI, Gordon Personality Inventory, Adult Personality Inventory, Maudsley Personality Inventory, Eysenck Personality Inventory, and Personality Research Form. Other social desirability scales included were Marlowe–Crowne Social Desirability Scale, Wiggins SD Scale, and Balanced Inventory of Socially Desirable Responding.

The search was restricted to published sources: English-language books, journals, and technical reports. Studies employing patient samples or children as participants were excluded, as our focus was on the normal population of individuals who are likely to be job applicants in the workforce.

The studies included in the meta-analyses were read and coded. In addition to the correlations, the scales used, the reliabilities reported, and sample characteristics were also coded. Psychometric meta-analysis (Hunter & Schmidt, 1990) was used to cumulate the results across studies and estimate the correlations of interest. In addition to sampling error, corrections were made for unreliability in the measures. Artifact distributions were used to correct for unreliability in the measures as the information was not available to correct each study individually. No corrections were made for range restriction, as this statistical artifact has been shown to be inconsequential for personality variables used in personnel selection (Barrick & Mount, 1991; Ones et al., 1993). The interactive procedure incorporating the meta-analytic refinements (e.g., use of mean correlation in sampling error formula) was used.

An artifact distribution was constructed for social desirability scales. Coefficient alphas and test–retest reliabilities over short time periods were included in the artifact distribution. Across 119 reliabilities, the mean estimate of social desirability scale reliability was .74 and the associated standard deviation was .14.

Based on the means and standard deviations reported for men and women, standardized differences (also referred as *d* values or effect sizes) between the two groups in terms of social desirability were computed. The meta-analysis cumulated these effect sizes. The results of this meta-analysis are reported in Table 3. The meta-analysis of the social desirability–age relation cumulated correlations. These results are reported in Table 4.

TABLE 3
Meta-Analysis of the Relation Between Social Desirability Scale Scores and Gender

Demographic Variable	*K*	*N*	*Mean d*	Var_{obs}	*Corrected d*
Sex	66	17,906	−.19	.0613	−.22

Note. *K* = number of effect sizes; Mean *d* = mean observed effect size (*d*); Var_{obs} = sample size weighted observed variance; Corrected *d* = relation corrected for unreliability in social desirability measures.

TABLE 4
Meta-Analysis of the Relation Between Social Desirability Scale Scores and Age

Demographic Variable	*K*	*N*	*Mean r*	Var_{obs}	*Corrected r*
Age	19	4,594	.10	.0225	.12

Note. K = number of correlations; Mean r = mean observed correlation (r); Var_{obs} = sample size weighted observed variance; Corrected r = relation corrected for unreliability in social desirability measures.

Across 66 studies (N = 17,906) results indicate that the sample size weighted observed mean d value was –.19. Across these studies, men scored on average .19 standard deviation units higher than women. The corrected sample size weighted d value (corrected for unreliability in the social desirability measures) was –.22. Cohen (1977) identified effect sizes of .20 or less as small. Thus, gender differences in social desirability appear to be small.

Across 19 studies (N = 4,594), results indicate that sample size weighted observed correlation between age and social desirability was .10. The corrected sample size weighted correlation (corrected for unreliability in the social desirability measures) was .12. It appears that social desirability displays consistent but mild correlations with age. Older individuals appear to score somewhat higher on social desirability scales.

The findings from the meta-analyses indicate that men score somewhat higher on social desirability scales. Similarly, older individuals tend to score higher on these scales, even though the relative effect is smaller. The practice of adjusting substantive personality scale scores on the basis of responses on social desirability scales may be expected to result in slightly larger adjustments for men and older individuals. However, it is worth pointing out that social desirability scales are not likely to be the cause of any adverse impact for women and older individuals.

SOCIAL DESIRABILITY AND CRITERION-RELATED VALIDITY

One question of paramount importance involves the influences of social desirability on criterion-related validity. Recently, arguments have been made that criterion-related validity is not affected by faking, but fakers will rise to the top. This is perplexing. Criterion-related validity is a beacon that has guided the science of industrial psychology and personnel selection research. It is an actuarial index of how well a predictor functions (Nunnally, 1978). The legal system and professional standards rely on criterion-related validity in establishing job-relatedness (Cascio, 1991; Schneider & Schmitt, 1986). Those who ignore criterion-related validity in

demonstrating the usefulness and job relatedness of selection measures do so at their own peril (Hogan & Hogan, in press). It is also important to note that the employer interested in selecting a predictor is concerned with the question of whether that predictor distinguishes good from poor performers in the population of potential *job applicants*, and not whether the relation between test scores and performance holds for *those hired* (Gulliksen, 1950; Thorndike, 1971). To argue that a selection instrument has predictive validity but fakers rise to the top is a rejection of the actuarial model that is at the heart of industrial psychology in favor of a clinical approach to selection. Calls to do away with predictive validity when evaluating the usefulness of personality scales is also equivalent to shifting into a clinical approach in personnel selection. This, we reject. Demonstrating criterion-related validity is crucial for all predictors and therefore examining social desirability influences on criterion-related validities of personality and integrity tests is critical.

There are probably four potential hypotheses regarding the role of social desirability in using personality scales and integrity tests in personnel selection. Social desirability can function as (a) a predictor, (b) a moderator, (c) a mediator, or (d) a suppressor variable. These four roles cover all the different roles that social desirability can play in personnel selection situations. We first discuss empirical evidence for social desirability as a predictor. Then we empirically evaluate the moderator, mediator, and suppressor hypotheses for personality scales and integrity tests.

Social Desirability as a Predictor of On-The-Job Behaviors

First, does social desirability function as a predictor of work attitudes and behaviors? The rationale behind this conceptualization is that ability to self-enhance may be regarded as an aspect of social competence. Those job applicants who are able to distort their responses in a socially desirable direction may be the same individuals who are able to be successful in interpersonal interactions at work. If substantial correlations are found between social desirability scales and external criteria, it will lead to the conclusion that social desirability is not a response bias, but a predictor in its own right. This is the case where social desirability as a contaminating bias in personality is correlated with the criterion. The consequence of this finding is that even if the estimated population validity of a personality measure is zero, the observed validity would be positive due to the criterion-correlated contaminating influence of social desirability. That is, the result of criterion-correlated contamination on the validities of focal personality variables is to artificially increase the criterion-related validity. This conceptualization of social desirability's role in prediction posits that socially desirable responding may contribute to the prediction of job performance (Cohen & Lefkowitz, 1974; Nicholson & Hogan, 1990).

Moorman and Podsakoff (1992) presented a meta-analysis of social desirability correlations with organizational behavior variables. Some of their results are presented in Table 5. In the Moorman and Podsakoff meta-analysis, observed

correlations were corrected for unreliability in the measures. Social desirability correlated .22 with job satisfaction and .18 with organizational commitment. Interestingly, social desirability correlated –.18 with role conflict and –.24 with role ambiguity.

One concern that may arise here is the self-report nature of all the variables being correlated. Does social desirability correlate with externally measured criteria? Ones et al. (1996) meta-analytically examined this question. The external criteria examined were: school success, task performance, training performance, counterproductive behaviors, and job performance. The operational validities of social desirability scales for predicting external job relevant criteria are presented in Table 6. Note that the operational validities presented have been corrected for sampling error and unreliability in the criterion alone.

TABLE 5
Correlations of Social Desirability With Organizational Behavior Variables

Correlate	*K*	*N*	*r*	*Corrected r*	*Variance of Corrected r*	*95% Credibility Interval*
Job satisfaction	6	3,361	.17	.22	.0146	–.02 to .46
Satisfaction with supervisor	5	2,599	.02	.03	.0188	.24 to .30
Role conflict	4	2,182	–.13	–.18	.0108	–.38 to .02
Role ambiguity	6	2,599	–.17	–.24	.0031	–.35 to –.13
Organizational commitment	5	5,506	.15	.18	.0023	.09 to .27
Performance	7	2,392	.01	.01	.0117	–.20 to .22

Note. Data selectively summarized from "A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational research," by R. H. Morrman and P. M. Podsakoff, 1992, *Journal of Occupational and Organizational Psychology, 65,* p. 136. Copyright 1996 by British Psychological Society. Adapted with permission.

TABLE 6
Correlations of Social Desirability With External Job-Related Criteria

Criterion	*K*	*N*	r_{mean} *With Social Desirability*	*Operational Validity of Social Desirability*
School success	16	3,125	–.09	–.11
Task performance	6	3,230	.00	.00
Training performance[a]	7	4,547	.19	.22
Counterproductive behaviors	6	1,479	–.03	–.03
Job performance[b]	14	9,966	.01	.01

Note. From "The role of social desirability in personality testing for personnel selection: The red herring," by D. S. Ones, C. Viswesvaran, and A. D. Reiss, 1996, *Journal of Applied Psychology, 81,* pp. 660–679. Copyright 1996 by American Psychological Association. Reprinted with permission.

[a]Instructor ratings of training performance. [b]Supervisory ratings.

The meta-analysis of 16 validities (N = 3,125) for predicting school success (grade point average [GPA] or course grades) indicated that the operational validity of social desirability scales for this particular criterion was –.11. The operational validity for task performance was .00 (N = 3,230). However, social desirability scales predicted training performance with an operational validity of .22 (N = 4,547). The operational validity of social desirability scales for predicting counterproductive behaviors is –.03 (N = 1,479). Finally, the operational validity of social desirability scales for predicting supervisory ratings of job performance was found to be .01 (N = 9,966).

One interesting question involves the relations among social desirability and dimensions of job performance at a more fine-grained level. Hough et al. (1990) reported the observed correlations between social desirability and five job performance dimensions using Project A data. These performance dimensions were: technical proficiency, general soldiering proficiency, effort and leadership, personal discipline, physical fitness, and military bearing. The observed correlations for these performance dimensions were –.07, –.06, .02, .05, .07, respectively. Hough et al. (1990) findings are summarized in Table 7.

Taken together, results from Moorman and Podsakoff (1992), Ones et al. (1996), and Hough et al. (1990) indicate that social desirability may be a useful predictor of job satisfaction, organizational commitment, and ratings of training performance. However, correlations with other externally measured performance criteria are small enough to be considered negligible.

Social Desirability as a Moderator, Mediator, or Suppressor

Next we review the influence of social desirability on personality scale validities. We examine moderator, mediator, and suppression hypotheses. Social desirability

TABLE 7
Criterion-Related Validities of ABLE Response Validity Scales

Response Validity Scale	*Technical Proficiency*	*General Soldiering Proficiency*	*Effort and Leadership*	*Personal Discipline*	*Physical Fitness and Military Bearing*
Nonrandom Response	.13	.14	.07	.10	.02
Social Desirability	–.07	–.06	.02	.05	.07
Poor Impression	–.04	–.05	–.15	–.15	–.16
Self-Knowledge	–.04	–.03	.07	.05	.13

Note. Data selectively summarized from 'Criterion-related validities of personality constructs and the effect of response distortion on those validities," by L. M. Hough, N. K. Eaton, M. D. Dunnette, J. D. Kamp, and R. A. McCloy, 1990, *Journal of Applied Psychology, 75,* pp. 581–595. Copyright 1990 by American Psychological Association. Adapted with permission. N's ranging between 7,666 to 8,477.

can be construed as a moderator if personality scale validities are different as a function of social desirability (i.e., less valid among high faking groups). If social desirability is a mediator of personality–performance relations, it should correlate with the criterion as well as the personality variables. If social desirability is a suppressor, it should correlate with the personality variables but not with the criterion. In examining these hypotheses, we take up personality scales first and then turn our attention to integrity tests.

Personality scales. There have been four large-scale meta-analyses of personality variables used in the prediction of job performance and its dimensions (Barrick & Mount, 1991; Hough et al., 1990; Salgado, 1997; Tett et al., 1991). All these meta-analyses concluded that personality measures could fruitfully be employed in personnel selection. Moreover, Hough et al. specifically examined the criterion-related validities of personality scales and the effect of response distortion on those validities. To examine whether nonrandom response, social desirability, poor impression, and self-knowledge (the four response validity scales) moderated the validities of the ABLE content scales, Hough et al. conducted a series of analyses (note that for these analyses only three criteria were used). Their results are summarized in Table 8.

To examine whether social desirability moderated the criterion-related validities of personality scales, Hough et al. (1990) divided their sample to two groups: an "overly desirable responding" group (*N*s = 2,428 to 2,480) and an "accurately responding" group (*N*s = 5,896 to 5,997). They computed the criterion-related validities of the 11 ABLE content scales separately for these two groups. Their conclusion was that the criterion-related validities for the "overly desirable responding" group were not substantially different from the validities obtained for the "accurately responding" group. Note that the criterion-related validities of personality variables are similar for accurately responding and overly desirable responding soldiers. Based on these results, Hough et al. concluded that social desirability did not moderate the validities. Hough et al. stated that personality scales could fruitfully be employed in personnel selection, and that "social desirability may not be the problem it has often assumed to be" (p. 592).

How about mediation or suppression hypotheses? Could social desirability be a mediator or a suppressor variable? Because there are negligible relations between social desirability and job performance, social desirability cannot play a mediator role in personnel selection systems using personality measures (Ones et al., 1996).

Social desirability scales may not correlate with job performance, but may be related to real individual differences in personality. This is the case where we have a variable uncorrelated with the criterion, but correlated with the predictor—a suppressor variable. This conceptualization of the role of social desirability in predicting external criteria points to it as a suppressor variable that does not correlate with the criterion of interest, but which through its correlation with the predictor

TABLE 8
Effects of Social Desirability on Criterion-Related Validities of Assessment of Background and Life Experience (ABLE) Scales: Testing the Moderator Hypothesis

	Effort and Leadership		*Personal Discipline*		*Physical Fitness*	
Scale	*Accurate*	*Overly Desirable*	*Accurate*	*Overly Desirable*	*Accurate*	*Overly Desirable*
Surgency						
Dominance	.15	.14	.00	.06	.18	.17
Energy Level	.23	.20	.13	.15	.27	.20
Achievement						
Self-Esteem	.21	.18	.12	.12	.21	.17
Work Orientation	.25	.20	.17	.16	.22	.17
Adjustment						
Emotional Stability	.17	.16	.11	.12	.16	.13
Agreeableness						
Cooperativeness	.16	.13	.20	.21	.14	.12
Dependability						
Traditional Values	.14	.11	.26	.22	.18	.11
Nondelinquency	.13	.12	.28	.29	.14	.11
Conscientiousness	.19	.14	.22	.22	.24	.14
Locus of Control						
Internal Control	.13	.12	.12	.15	.15	.08
Physical Condition						
Physical Condition	.08	.09	−.03	.02	.28	.29

Note. Data selectively summarized from 'Criterion-related validities of personality constructs and the effect of response distortion on those validities," by L. M. Hough, N. K. Eaton, M. D. Dunnette, J. D. Kamp, and R. A. McCloy, 1990, Journal of Applied Psychology, 75, pp. 581–595. Copyright 1990 by American Psychological Association. Adapted with permission. Overly desirable group N = 2,428–2,480; accurate group N = 5,896–5,997.

suppresses invalid personality scale variance (Dicken, 1963; Ganster, Henessey, & Luthans, 1983; Ruch & Ruch, 1967). Ones et al. (1996) also examined the influence of social desirability on personality variable criterion-related validities. Table 9 summarizes the Ones et al. (1996) results.

The results in Table 9 suggest that partialling social desirability from personality measures does not have any impact on the criterion-related validities of the Big Five variables. The partialling process leaves the validities intact. From these results, it appears that social desirability does not attenuate the criterion-related validities of personality dimensions. As such, social desirability does not appear to be a response bias that attenuates criterion-related validity of personality variables for job performance. Ones et al. (1996) concluded that even though social desirability is in a position to explain unique variance in conscientiousness, is unrelated to job performance, and therefore is poised to function as a suppressor, it has little importance as such.

Integrity tests. Ones et al. (1993) reported a meta-analysis of integrity test validities (based on 665 validity coefficients across 576,460 data points) for three criterion categories: (a) job performance, (b) general counterproductive behaviors at work, and (c) theft. Key results from Ones et al. that bear on the social desirability question are summarized in Table 10.

In selection settings, the best estimate of integrity test validities for predicting job performance is based on (a) predictive studies (b) conducted on samples of applicants. To obtain such an estimate of the mean validity of integrity tests for selection, Ones et al. (1993) meta-analyzed predictive validities calculated on applicant samples. There were 23 such validities for predicting supervisory ratings of job performance. Across 7,550 people, the best estimate of the mean true validity

TABLE 9
Influence of Social Desirability on Personality Variable Validities for Various Criteria: Testing the Suppression Hypothesis

Personality Dimension	ρ	ρ *After Partialling Social Desirability*
Emotional Stability	.07	.07
Extraversion	.10	.10
Openness to Experience	−.03	−.03
Agreeableness	.06	.06
Conscientiousness	.23	.23

Note. From "The role of social desirability in personality testing for personnel selection: The red herring" by D. S. Ones, C. Viswesvaran, and A. D. Reiss, 1996, *Journal of Applied Psychology, 81,* pp. 660–679. Copyright 1996 by American Psychological Association. Reprinted with permission.

TABLE 10
Validity of Integrity Tests for Supervisory Ratings of Overall Job Performance: Testing the Moderator Hypothesis

Validation Strategy	*Sample*	*N*	*K*	*Mean r*	ρ	SD_ρ	*% Var. Acc.*	*90% CV*
Predictive	Applicants	7,550	23	.25	.41	.00	100	.41
Predictive	Employees	8,994	20	.15	.26	.21	24.4	.01
Concurrent	Employees	8,275	63	.22	.37	.14	61.0	.21

Note. Data selectively summarized from Table 8 of "Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]," by D. S. Ones, C. Viswesvaran, and F. L. Schmidt, 1993, *Journal of Applied Psychology, 78,* pp. 679–703. Copyright 1993 by American Psychological Association. Adapted with permission. Table summarizes information across overt and personality-based integrity tests. K = number of correlations; Mean r = mean observed correlation; r = operational validity (mean r corrected for range restriction and unreliability in the criterion only); SD_r = standard deviation of the operational validity; % Var. Acc. = percentage variance due to all corrected statistical artifacts; 90% CV = lower 90% credibility value.

was .41. This is the validity of integrity tests for job applicant populations (all studies contributing to this analysis were conducted on job applicants). The standard deviation of true validity was 0, since the proportion of the observed variance accounted for by artifactual variance was 100%. These findings imply that the average validity integrity tests may be expected to have in selection settings for supervisory ratings of overall job performance is .41, and that this value is *constant* across settings. Thus, differential social desirability influences across settings do not moderate the criterion-related validities of integrity tests.

With integrity tests it is critical to note that the criterion-related validity evidence is based mostly on job applicant samples. In the Ones et al. (1993) meta-analysis, for example, any response distortion that may have been engaged in by job applicants still left the criterion-related validity at .41 for supervisory ratings of job performance, and at .32 for general counterproductive behaviors on the job. The criterion-related validities of integrity tests are substantial even for job applicant samples.

The mediation and suppression-based influences of social desirability on the criterion-related validities of integrity tests have not been studied to date. In our recent work (Ones & Viswesvaran, in press), we conducted a meta-analysis examining the relation between social desirability and integrity. The PsycLIT database was searched for studies reporting on integrity and social desirability for the time period 1974 to 1995. A manual search of the psychological abstracts was also undertaken, covering the years 1945 to 1973. A snowballing technique, whereby the references of the obtained articles were searched to identify further relevant studies, was also used. Twenty studies were identified from the published literature, reporting a correlation between an integrity test and a social desirability scale. The studies were read and coded. Psychometric meta-analysis (Hunter & Schmidt, 1990) was used to cumulate the results across studies. In addition to sampling error, corrections were made for unreliability in the measures. The integrity test unreliability artifact distribution was the one used by Ones et al. (1993). The social desirability scale unreliability artifact distribution was the one used by Ones et al. (1996). Results are presented in Table 11.

TABLE 11
Meta-Analysis of the Relation Between Integrity Test Scores and Social Desirability

N	*K*	*Mean r*	Var_{obs}	*Sampling Error of the Mean*	*90% Confidence Interval*	ρ *Between Integrity and Social Desirability*
3,973	20	.06	.02	.03	.02 to .10	.08

Note. Meta-analysis was based on published studies only. *K* = number of correlations being cumulated; Mean *r* = mean observed correlation; Var_{obs} = sample size weighted observed variance; Sampling error of the mean = estimated sampling error of the mean observed correlation; 90% Confidence Interval = 90% confidence interval constructed around the mean observed correlation; ρ = estimated population correlation (corrected for unreliability in both measures).

The estimated true score correlation between social desirability and integrity constructs was .08 (N = 3,973). The observed (mean uncorrected) correlation between the measures of integrity and social desirability was .06 (observed SD = .02). The 90% confidence interval around the mean observed correlation ranged from .02 to .10. These results suggest that there is negligible overlap between the constructs measured by integrity tests and social desirability scales.

Obtaining a meta-analytic estimate of the correlation between integrity and social desirability also facilitates a test of several process mechanisms by which social desirability can affect integrity test–criterion relations. We examined the effect of controlling for social desirability on integrity test validities. The three criteria for which these investigations could be carried out were: training performance, counterproductive behaviors, and job performance. The inputs to these investigations were the social desirability–criterion, integrity–criterion, and integrity–social desirability correlations. The social desirability–criterion correlations were obtained from Ones et al. (1996). The integrity test correlations with counterproductive behaviors and job performance were taken from Ones et al. (1993). Integrity test correlations with training performance and social desirability were taken from analyses reported in Ones and Viswesvaran (in press). The effects of partialling social desirability from integrity test criterion-related validities are presented in Table 12.

TABLE 12
Influence of Social Desirability on Integrity Test Validities for Various Criteria: Testing the Suppression Hypothesis

Criterion	*Operational Validity of of Social Desirability Scales*	*Operational Validity of Integrity Tests*	*Operational Validity of Integrity Tests After Partialling Social Desirability*
Training performance	.22	.38	.36
Counterproductive behaviors	−.03	.32	.32
Job performance	.01	.41	.41

Note. Operational validity of social desirability scales summarized from meta-analysis presented in "The role of social desirability in personality testing for personnel selection: The red herring," by D. S. Ones, C. Viswesvaran, and A. D. Reiss, 1996, *Journal of Applied Psychology, 81,* pp. 660–679. Copyright 1996 by American Psychological Association. Adapted with permission. Operational validity of integrity tests summarized from meta-analysis presented in "Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]," by D. S. Ones, C. Viswesvaran, and F. L. Schmidt, 1993, *Journal of Applied Psychology, 78,* pp. 679–703. Copyright 1993 by American Psychological Association. Adapted with permission. Operational validity of integrity tests after partialling social desirability has been computed by using the correlation between social desirability scales and integrity tests reported in Table 11 of this article.

With regard to each of the three criteria (training performance, counterproductive behaviors, and job performance), three hypotheses can be tested using the partial correlations reported in Table 12. First, if partialling social desirability increases validities of integrity tests for the criteria, then the conclusion that social desirability is a suppressor variable may be reached. The implication of this conclusion would be to control for social desirability in research and practice. Second, if partialling social desirability reduces validities of integrity tests for various criteria, then the conclusion is that social desirability contributes to the prediction of the criteria. The implication of this conclusion would be to include social desirability scales in predictor batteries along with integrity tests. Third, if partialling social desirability leaves the integrity test validities intact, one can conclude that social desirability is irrelevant as a suppressor or as having potential in contributing to the prediction of criteria.

Results presented in Table 12 indicate that partialling social desirability from integrity test validities for counterproductive behaviors and job performance results in no change in the criterion-related validities. Social desirability is not a worthwhile predictor or suppressor variable with regard to these two criterion variables. Partialling social desirability from integrity test validities for training performance reduced the validity if integrity tests from .38 to .36, indicating some independent criterion-related validity for the social desirability construct. However, the incremental validity of social desirability scales over integrity tests is small enough to be practically unimportant.

In general, the results presented in Tables 10 and 12, coupled with findings from Ones et al. (1996), point out that social desirability is not (a) a strong predictor of job-related criteria, (b) a mediator of integrity–criterion relations (for this social desirability would need to be related to both integrity test scores and criteria), and (c) a suppressor variable with regard to integrity test validities. Similar to results for personality scales, integrity tests' criterion-related validities are unaffected by social desirability.

A Theoretical Explanation of Why Social Desirability Does Not Influence Criterion-Related Validity

That socially desirable responding does not destroy correlations with external criteria is also founded on a theory of personality in general, and a theory of personality item responses in particular (Hogan & Hogan, in press). This theory postulates a model of individual responses to items comprising personality scales that explains why the criterion-related validity may not be affected by socially desirable responding. The socioanalytic theory of personality (Hogan & Hogan, in press) construes an individual's responses to items as a social interaction between the test taker (job applicant) and the test administrator (employer). The individual

responds to items so as to negotiate an identity for oneself. The individual would also behave, to the extent possible, to be consistent with that identity. Thus, an individual who is motivated to negotiate an identity of a conscientious individual on personality scales is also likely to strive for that identity by engaging in trustworthy and dependable behaviors. Therefore, if employers seek to hire individuals who will behave conscientiously on the job (or at least seek to project a conscientious image to others, especially supervisors), selecting individuals who respond to items on a personality scale so as to project an identity of a conscientious individual is sensible. According to Hogan and Hogan (in press), responding to personality items during pre-employment testing is one way of "telling others" how one wants to be seen, whereas everyday social interactions on the job is another way.

DISCUSSION AND CONCLUSIONS

A common concern expressed about the use of personality scales and integrity tests in personnel selection goes as follows: "Honest individuals who do not fake will be rejected (precisely those whom the organization aims to select with integrity test use) whereas those who fake their responses will get selected (precisely those who should be screened out)." Although intuitively appealing, this argument is not supported by the available data. Large-scale meta-analyses have shown that faking does not destroy predictive validity in the personality domain in general (cf. Hough et al., 1990; Ones et al., 1996) and integrity tests in particular (cf. Ones et al., 1993). This is a conclusion strengthened by the analyses summarized here as well.

In asserting that social desirability does not affect the usefulness of personality measurement for personnel selection, we want to make clear that all steps should be taken to standardize the test administration process and to engender a common frame of reference to all job applicants. Both these are important due to their direct influence on reliability. That is, one may anticipate the reliability to be enhanced when a common frame of reference is used across job applicants. A good example of this mechanism was illustrated by Schmit, Ryan, Stierwalt, and Powell (1995). Schmit et al. increased the shared frame of reference among test takers (students) by contextualizing a conscientiousness measure. This was done by adding "at school" tags to all personality items. The contextualized conscientiousness measure produced a validity of .41 in predicting GPA. When students completed the contextualized scale as though they were applying for college admission, the validity for GPA was found to be .46. In both cases, these validities were higher than those found for the noncontextualized version of the conscientiousness measure. It appears that reliability improvements afforded by imparting a common frame of reference among test takers improves validity as well.

Another issue to consider is that the issue of faking or response distortion is a concern for all selection instruments (be they interviews or assessment centers).

Raising faking as a concern in using personality scales and integrity tests for personnel selection inherently suggests a concern that faking is more prevalent or more critical in using personality measures and integrity tests than it is in using other predictors (interviews, interest inventories, assessment centers, biodata, etc.). There simply is no empirical evidence to substantiate this assertion that faking is more prevalent in personality assessment and integrity tests than in other predictors. Among the critics, there exists only an assumption that faking is more likely for personality measures and integrity tests than it is in other predictors such as interviews.

Despite claims from critics of noncognitive measures used in personnel selection, real-world data show that social desirability is not a factor destroying the criterion-related validity of personality measures and integrity tests. Data from applicants and from large-scale meta-analyses indicate that faking does not matter in prediction for personnel selection. Our earlier work identified social desirability as the red herring in personality measurement. Our data from real-world job applicants confirm that criticizing personality scales because of potential response distortion by applicants is making much ado about nothing. The predictive success of personality and integrity measures in the prediction of job performance and its dimensions might be a hard pill to swallow for those who do not believe that noncognitive measures have their place in personnel selection systems. Personality scales and integrity tests have incremental validity over cognitive measures and they add this incremental validity while decreasing adverse impact (see Hough, in press; Ones et al., 1993; Ones & Viswesvaran, 1996; Ones & Viswesvaran, 1998). But condemning personality measurement and integrity testing because of unproven potential social desirability problems is beating a dead horse. We suggest that the massive real-world data from applicant samples have nailed the coffin of social desirability and faking shut. Integrity tests and personality measures have proven their worth in personnel selection systems (e.g., Barrick & Mount, 1991; Costa, 1996; Hogan & Hogan, 1989; Hough et al., 1990; Hough, 1992, in press; Ones et al., 1993; Ones et al., 1996; Sackett & Wanek, 1996; Schmit et al., 1995; Tett et al., 1991). Those who are threatened by such measures should find a scapegoat other than the unrealized potential of social desirability influences on criterion-related validity of personality and integrity measures used in personnel selection.

ACKNOWLEDGMENTS

We thank Lyn Avey, Nathan Kuncel, Chi Ngo, and Alicia Stevens for help with article collection, organization, and manuscript preparation phases of this article.

The order of authorship is arbitrary, the two authors contributed equally to this article.

Portions of this article were presented at the 1997 annual conferences of the Society for Industrial and Organizational Psychology, St. Louis, MO and American Psychological Association, Chicago.

REFERENCES

Barrick, M. R. , & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI.* New York: Appleton-Century Crofts.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for the restandardized Minnesota Multiphasic Personality Inventory: MMPI–2. An Administrative and interpretative guide.* Minneapolis: University of Minnesota Press.

Campbell, D. T. , & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cascio, W. F. (1991). *Applied psychology in personnel management* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Cellar, D. F., Miller, M. L., Doverspike, D. D., & Klawsky, J. D. (1996). Comparison of factor structures and criterion-related validity coefficients for two measures of personality based on the five factor model. *Journal of Applied Psychology, 81,* 694–704.

Cohen, J. (1977). *Statistical power analysis for the behavior sciences* (Rev. ed.). New York: Academic.

Cohen, J., & Lefkowitz, J. (1974). Development of a biographical inventory blank to predict faking on personality tests. *Journal of Applied Psychology, 59,* 404–405.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Costa, P. T. (1996). Work and personality: Use of the NEO–PI–R in industrial/organizational psychology. *Applied Psychology: An International Review, 45,* 225–241.

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory manual.* Odessa, FL: Psychological Assessment Resources.

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

Dean, C. C. (1990). *'Paper and Pencil' honesty tests: A study for determining their susceptibility to faking.* Unpublished manuscript, University of North Texas.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement, 23,* 699–720.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7,* 385–400.

Gannon, L., Raber, A., Jenkins, M., Ketterman, R., & Griffith, R. (1997, Spring). *Sex and gender role differences in the faking of non-cognitive selection measures.* Presented at the Organizational Behavior Graduate Conference.

Ganster, D. C., Henessey, H. W., & Luthans, F. (1983). Social desirability response effects: Three alternative models. *Academy of Management Journal, 26,* 321–331.

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

Gough, H. G., & Heilbrun, A. B., Jr. (1983). *The Adjective Checklist Manual: 1983 Edition.* Palo Alto, CA: Consulting Psychologists Press.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Hamill, L. S., & Wheeler, J. K. (1997, April). *Gender differences in "Faking Good" on Personality Tests.* Symposium presented at the annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Hogan, J., & Hogan, R. (1989). How to measure employee reliability. *Journal of Applied Psychology, 74,* 273–279.

Hogan, J., & Hogan, R. (1991, April). *Levels of analysis in the big five theory: The structure of self-description.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Hogan, R., & Hogan, J. (1995). *Hogan personality inventory manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.

Hogan, J., & Hogan, R. (in press). Theoretical frameworks for assessment. In P. R. Jeanneret & R. Silzer (Eds.), *Individual psychological assessment.* San Francisco: Jossey-Bass.

Hogan, R., & Jones, W. (1992). *Interpretative manual for the inventory of personal motives.* Tulsa: Hogan Assessment Systems.

Hough, L. M. (1992). The "Big Five" personality variables–construct confusion: Description versus prediction. *Human Performance, 5,* 139–155.

Hough, L. M. (in press). The millennium for personality psychology: New horizons or good old daze. *Applied Psychology: An International Review.*

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75,* 581–595.

Hough, L. M., & Schneider, R. J. (1996). The frontiers of I/O personality research. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 31–88). San Francisco: Jossey-Bass.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Jones, W. (1990). *Interpretive manual for the PROFILE.* Tulsa: Hogan Assessment Systems.

Kelly, E. L., Miles, C. C., & Terman, L. M. (1936). Ability to influence one's score on a typical pencil-and-paper test of personality. *Character and Personality, 4,* 206–215.

Livneh, H., & Livneh, C. (1989). The five-factor model of personality. Is evidence of its cross-measure validity premature? *Personality and Individual Differences, 10,* 75–80.

LoBello, S. G., & Sims, B. N. (1990). Fakability of a commercially produced pre-employment integrity test. *Journal of Business and Psychology, 8,* 265–273.

McCrae, R. R., & Costa, P. T., Jr. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

Michaelis, W., & Eysenck, H. J. (1971). The determination of personality inventory factor patterns and intercorrelations by changes in real-life motivation. *The Journal of Genetic Psychology, 118,* 223–234.

Moore, R. W. (1990) Instructional effects on the Phase II profile honesty test. *Psychological Reports, 67,* 291–294.

Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behavior research. *Journal of Occupational and Organizational Psychology, 65,* 131–149.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and applications.* Upper Saddle River, NJ: Prentice Hall.

Nicholson, R. A., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist, 45,* 290–292.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Ones, D. S. (1993). *The construct validity of integrity tests.* Unpublished doctoral dissertation, University of Iowa.

Ones, D. S., Schmidt, F. L., & Viswesvaran, C. (1994, August). Intercorrelations among the Big Five personality scales: Implications for personnel selection. In R. T. Hogan (Chair), *Personality Assessment in Personnel Selection*. Symposium conducted at the annual meeting of the American Psychological Association, Los Angeles.

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17,* 609–626.

Ones, D. S., & Viswesvaran, C. (1998). Gender, age and race differences on overt integrity tests: Analyses across four large-scale applicant data sets. *Journal of Applied Psychology, 83,* 35–42.

Ones, D. S., & Viswesvaran, C. (in press). Integrity testing in organizations. In R. W. Griffin, A. O'Leary-Kelly, & J. M. Collins (Eds.), *Dysfunctional behavior in organizations: Vol. 2. Nonviolent behaviors in organizations*. Greenwich, CT: JAI.

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

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance [Monograph]. *Journal of Applied Psychology, 78,* 679–703.

Ruch, F. L., & Ruch, W. W. (1967). The K factor as a (validity) suppressor variable in predicting success in selling. *Journal of Applied Psychology, 51,* 201–204.

Ryan, A. M., & Sackett, P. R. (1987) Preemployment integrity testing: fakability, reactions of test takers, and company image, *Journal of Business & Psychology, 1,* 248–256.

Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49,* 787–829.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82(1),* 30–43.

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966–974.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80,* 607–620.

Schneider, B., & Schmitt, N. (1986). *Staffing organizations* (2nd ed.). Prospect Heights, IL: Waveland.

Schwab, D. P. (1971). Issues in response distortion studies of personality inventories: A critique and replicated study. *Personnel Psychology, 24,* 637–647.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703–742.

Thorndike, R. L. (1971). *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology, 48,* 865–885.

Viswesvaran, C., & Ones, D. S. (in press). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement.*

Wanek, J. E. (1991). *Testing the fakability of the Personnel Reaction Blank relative to the fakability of an overt integrity test.* Unpublished Master's thesis, University of Minnesota, Minneapolis, MN.

Wiggins, J. S. (1991). *Manual for the Interpersonal Adjective Scales.* Odessa, FL; Psychological Assessment Resources.