The Impact of Response Distortion on Preemployment Personality Testing and Hiring Decisions

Joseph G. Rosse, Mary D. Stecher,
and Janice L. Miller
University of Colorado at Boulder

Robert A. Levin
Center for Human Function & Work

Response distortion (RD), or faking, among job applicants completing personality inventories has been a concern for selection specialists. In a field study using the NEO Personality Inventory, Revised, the authors show that RD is significantly greater among job applicants than among job incumbents, that there are significant individual differences in RD, and that RD among job applicants can have a significant effect on who is hired. These results are discussed in the context of recent studies suggesting that RD has little effect on the predictive validity of personality inventories. The authors conclude that future research, rather than focusing on predictive validity, should focus instead on the effect of RD on construct validity and hiring decisions.

Personality assessment as a preemployment screening procedure is receiving renewed interest from researchers and practitioners. A number of quantitative reviews have demonstrated that personality inventories can be useful predictors of job performance, particularly if specific, job-relevant personality constructs are used to predict specific criteria (Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Ones, Viswesvaran, & Schmidt, 1993; Tett, Jackson, & Rothstein, 1991). These findings have led to a resurgence of interest in personality testing as an employee-selection tool.

Yet this trend is not without controversy. One major debate concerns the effect of response distortion on personality inventory scores. What is clear from the existing research is that people completing personality inventories

can inflate their scores if they want to. A number of studies have compared "fake good" to "answer honestly" conditions and found substantial differences in scores; Ones, Viswesvaran, and Korbin's (1995) meta-analysis of this literature reported that faking can increase scores by nearly one-half standard deviation. What has been less clear is whether actual applicants engage in these levels of response distortion, and if so, what effect this distortion has on the validity, utility, and fairness of preemployment personality assessments. The purpose of this article is to explore the extent of response distortion on personality inventory scores in an actual applicant-testing environment and its potential effect on which applicants get hired.

The Prevalence of Response Distortion in Employment Screening

In an organizational setting, assessment procedures may create the motivation, as well as the opportunity, to distort responses in order to create a favorable self-presentation and a favorable outcome (Villanova & Bernardin, 1991). This is particularly so when assessment occurs in a context with strong demand characteristics, such as when applying for a job (Bass, 1957; Christiansen, Goffin, Johnston, & Rothstein, 1994; Leary & Kowalski, 1990; McCrae & Costa, 1983; Paulhus, 1991b). Given the motivation to make a good impression, applicants are likely to want to convey an image that (a) reflects the self-concept but is biased in a positive direction, (b) matches perceived role demands, and (c) exhibits the attributes of the prototypic or ideal employee (Leary & Kowalski, 1990). For example, Paulhus and Bruce (1991) found that when their participants were instructed to distort their responses to match hypothetical job profiles, they were

quite successful in doing so. Furthermore, Schmit and Ryan (1993) found a large "ideal-employee" factor in responses to the NEO Personality Inventory (Costa & McCrae, 1992) that was not present in a sample of college students.

Transparent questions such as those included in many personality inventories make it easier to engage in response distortion (Alliger, Lilienfeld, & Mitchell, 1995; Furnham, 1986). Trait descriptors used in most personality inventories tend to be value laden, making the social desirability of endorsing items easy to discern. For example, four of the Big Five factors are represented by predominantly positive terms (e.g., "assertive," "verbal," "energetic," "bold," "active," and "daring" for Extraversion; "helpful," "cooperative," "sympathetic," "warm," "trustful," "considerate," and "pleasant" for Agreeableness; "organized," "thorough," "practical," "efficient," "careful," and "hardworking," for Conscientiousness; and "unconventional," "open to new ideas," "questioning," "curious," "creative," and "imaginative" for Openness to Experience; Lillibridge & Williams, 1992). Neuroticism, on the other hand, is represented by mostly negative terms (e.g., "anxious," "moody," "temperamental," "emotional," "nervous," and "depressed"). In addition to their general social desirability, many of the items on Big Five inventories have answers that are obviously "correct" when applying for a job (e.g., "I am a productive person," "I don't like to waste time."). Mahar, Colognon, and Duck (1995) found that job applicants are likely to answer questions in terms of their role expectations, a form of faking that may be difficult to detect with typical response-distortion scales (Kroger & Turnbull, 1975).

A final factor thought to contribute to response distortion is the nonverifiability of responses on personality inventories. Generally speaking, there is no way to verify applicants' assertions that they are planful in their work, enjoy being around others, or tend to view life with optimism. Studies of response distortion on application blanks have shown that dishonest responses are more likely on questions that are not objective and cannot be verified (Becker & Colquitt, 1992). As Fiske and Taylor (1991) noted, people tend to overstate their abilities unless they believe their actual abilities will be verified.

Personality testing thus provides an almost ideal setting for dissimulation: Job applicants are motivated to present themselves in the best possible light; transparency of items makes it possible to endorse items that will make them look good, and there is little apparent chance of being caught in a lie. Under these circumstances it would be surprising if most job applicants did *not* fake some of their answers. The utility of distorting responses on personality inventories during job application has been noted, and even advocated, since the use of inventories became popular for selection. For example, the early 1950s classic, *The Organization Man*, contains a critique of the transparency of selection instruments of the day, and a detailed instructional appendix entitled, "How to Cheat on Personality Tests" (Whyte, 1956).

Nevertheless, not all assessment specialists believe that response distortion is a problem. An alternative view states that although examinees can distort their responses if instructed to do so, most applicants do not actually do so. For example, the manual for the Hogan Personnel Selection Series states that "the base rate of faking during the job application process is virtually non-existent" (J. Hogan & Hogan, 1986, p. 20). As a result, many of the latest personality inventories designed to measure the five-factor model of personality do not include a measure of response distortion.

The conclusion that response distortion is rare in operational settings relies on studies finding that the scores of respondents who were told to respond as if they were applying for a job were very similar to those of respondents told to answer "honestly" (e.g., Ryan & Sackett, 1987). The strongest empirical support for this position is the large-sample study by Hough et al. (1990). However, because Hough et al.'s study dealt with military personnel who had completed the personality inventory after they had been sworn in, it is not evident that these results generalize to more typical applicant settings. Similarly, Ryan and Sackett's participants were students for whom personality test scores had little or no practical relevance. More recent studies using samples of actual applicants, in contrast, have indicated that job applicants' personality scores are higher than those of incumbents (Barrick & Mount, 1996; Hough, 1995, 1996; Hunt, Hansen, & Paajanen, 1996; White & Moss, 1995).

The Effect of Response Distortion on Hiring Decisions

A second debate concerns the effect of response distortion on the validity and usefulness of personality testing for hiring employees. According to the currently prevailing argument, even if response distortion occurs, it does not affect the predictive validity of personality inventories (R. Hogan, Hogan, & Roberts, 1996; Hough & Schneider, 1996). Several studies have shown that controlling for the effects of response distortion does not significantly increase the correlations between personality scores and criterion measures (Barrick & Mount, 1996; Christiansen et al., 1994; Dicken, 1963; Hough, 1995; Hough et al., 1990; McCrae & Costa, 1983). On the basis of a meta-analysis, Ones and her associates concluded that socially desirable responding is not a suppressor of the validity of the Big Five factors of personality for employment deci-

sions (Ones et al., 1995; Ones, Viswesvaran, & Reiss, 1996). Yet a number of studies—some of them too recent to be included in the meta-analysis—show that corrections for response distortion can improve criterion-related validities (Douglas, McDaniel, & Snell, 1996; Hunt et al., 1996; Kamp, cited in Hough, 1996; Paajanen, 1987).

There are both substantive and methodological reasons to question Ones et al.'s conclusion that response distortion is not a significant problem for personality testing. From a substantive perspective, the Ones et al. meta-analysis did not distinguish between two qualitatively distinct dimensions of socially desirable responding. Studies spanning the last three decades provide support for a two-factor structure of socially desirable responding (see Paulhus, 1991b, for a review). The first factor represents a form of unconscious ego-enhancement manifested by overly positive beliefs about the self-concept (Paulhus, 1991b; Sacheim & Gur, 1978). As such, this factor is thought to represent a dimension of personality in and of itself, labeled "self-deceptive positivity." Paulhus (1991b) reported substantial correlations between this self-deception factor and measures of adjustment; for example, high scorers on self-deception report high self-esteem and low levels of depression and neuroticism. Thus the self-deceptive factor appears to represent content variance in personality measures that should not be used as a control variable.

In contrast, the second factor of socially desirable responding represents deliberate tailoring of answers to create a positive impression, or what we refer to in this article as *response distortion* (Paulhus 1991b; Zerbe & Paulhus, 1987). Unlike scores on the first factor, scores on this dimension have been found to be particularly sensitive to situational demands (Paulhus & Reid, 1991). It is this intentional distortion that introduces construct-irrelevant variance that may affect the validity and utility of personality scale scores.

Unfortunately, most studies assessing the effects of response distortion have used measures that load on both dimensions of socially desirable responding, for example, the Marlowe-Crowne Social Desirability scale (Crowne & Marlowe, 1960), Minnesota Multiphasic Personality Inventory (MMPI; Dahlstrom, Welsh, & Dahlstrom, 1972), Lie scale, and the California Psychological Inventory Good Impression scale (Gough, 1975). A key contribution of this study is the use of a "pure" measure of response distortion, the so-called Impression Management scale of Paulhus' (1991a) Balanced Inventory of Desirable Responding.

There are also methodological reasons to believe that response distortion may have effects that have not been detected in prior studies. Dragow and Kang (1984) have shown that correlation coefficients are extremely robust estimators of linear associations between variables but

that this robustness comes at the cost of sensitivity to changes in rank order in particular ranges of a bivariate distribution (such as changes among the top-scoring applicants). Simply put, correlational analysis may be insensitive to changes in the rank ordering of applicants that are due to differences in response distortion. Although the observed (concurrent) validity of the test may not change for the whole sample, its validity for the applicants who are at the top end of the predictor distribution (corresponding to applicants who are most likely to be hired) may approach zero if response distortion occurs primarily among those who receive the highest scores (Douglas et al., 1996; Levin, 1995; Zickar, Rosse, & Levin, 1996).

There are four additional methodological factors that may reduce the sensitivity of correlation analysis in detecting effects of distortion: a skewed distribution of response distortion, selection ratio, restriction of range, and the modest validities of personality inventories. Due to the strong situational demands inherent in a job application setting, applicants are likely to engage in moderate to high impression management; very few are likely to engage in low impression management (Levin, 1995). This negative skew interferes with the ability to detect associations between response distortion and other variables. Low selection ratios, which are typical in many hiring situations, exacerbate the problem of a skewed distribution of response distortion by restricting the range of the personality and performance criterion measures. This restriction of range will also attenuate correlations between the response distortion measure and predictor (personality) and criterion (performance) measures. Moreover, Conger and Jackson (1972) have shown that suppression effects are extremely hard to detect when predictive validities are of the moderate to low magnitude common in employment settings.

In sum, the insensitivity of correlation coefficients, skewed distributions, low selection ratios, and modest predictive validities represent statistical artifacts that make it highly unlikely that suppressor effects of response-distortion measures will actually be detected in most data sets. What is critically important to note is that response distortion may have a dramatic effect on who is hired, even though it has no detectable effect on predictive validity. Let us assume that only 5% of applicants engage in extreme response distortion on a personality inventory. Because of personality scales' susceptibility to faking, these applicants will score higher than more candid applicants who have the same true scores on the personality dimensions. In fact, a substantial number of the top-scoring applicants may be people who consciously distorted their test answers and do not in fact have high true scores.

To date, few studies have considered the effect of response distortion on actual hiring decisions. Christiansen et al. (1994) reported that corrections to the Sixteen Personality Factor Questionaire (16PF; Cattell, Eber, & Tatsu-

oka, 1970) did not affect validity but did affect hiring decisions, especially when the selection ratio was less than 50%. Becker and Colquitt (1992) concluded that faking on a biodata instrument had little effect on who was hired, although they did find significantly higher scores among applicants who were actually hired for the job (using a selection ratio of 50%, which would minimize the effects of response distortion on hiring, as discussed later).

The Current Study

This study was designed with two purposes in mind. The first was to examine response distortion by actual job applicants in a realistic employment context when completing a personality inventory. As we have noted earlier, prior evidence has been mixed and many studies (e.g., Hough et al., 1990; LoBello & Sims, 1993; Ryan & Sackett, 1987) did not focus on actual job application settings. The impression-management literature indicates that situational demands are an important factor affecting impression-management behavior (cf. Leary & Kowalski, 1990). On the basis of this literature, we expect that the situational demands inherent in applying for a job will produce higher levels of response distortion among job applicants than among job incumbents. Job applicants in this study completed the personality inventory as part of the process of applying for a job; therefore, they should be motivated to present themselves in the most positive light possible (Bass, 1957; Christiansen et al., 1994; Leary & Kowalski, 1990; McCrae & Costa, 1983; Paulhus, 1991b). Job incumbents, on the other hand, were told that the results of the personality inventories would be used for research purposes only and would not be seen by their managers. In this situation, the confidentiality of the testing combined with the reduced motivation to present themselves in a positive light should reduce response distortion on the part of incumbents.

Hypothesis 1: Response-distortion scores will be higher among job applicants than among incumbents.

Although the situational demands of the application process should elevate response-distortion scores among job applicants as a whole (relative to incumbents), we also hypothesized that there would be substantial variation among job applicants (Leary & Kowalski, 1990; Paulhus, 1991b). Although some applicants will engage in little response distortion, we expect that most applicants will exhibit moderate to high levels, and a few will exhibit what Levin (1995) termed "extra-conventional" levels of extreme response distortion.

Hypothesis 2: Job applicants' response-distortion scores will show substantial variance and will be negatively skewed.

Response distortion is likely to vary, depending on what dimensions of personality are being measured (Hough, 1996). Because job applicants are attempting to make themselves as attractive as possible, they are more likely to describe themselves as well adjusted, dependable, and achievement oriented (Paulhus, Bruce, & Trapnell, 1995). Thus, we expected applicants' response-distortion scores to be most highly correlated with the Neuroticism and Conscientiousness dimensions of personality. Depending on their perception of the job requirements, applicants may also seek to describe themselves as agreeable and extroverted. However, because these are less "universal" descriptors of jobs, we expected their associations with response distortion to be less strong than those for Neuroticism and Conscientiousness. We expected Openness to Experience to be least related to job performance and therefore least affected by response distortion. Because incumbents have few situational demands to make a positive impression, they were not expected to distort their answers as much as would applicants.

Hypothesis 3a: Applicants' response-distortion scores will be most highly correlated with Neuroticism and Conscientiousness, moderately correlated with Agreeableness and Extraversion, and least correlated with Openness to Experience.
Hypothesis 3b: Job applicants will score higher than incumbents on Conscientiousness, Agreeableness, and Extraversion and score lower than incumbents on Neuroticism.

Our second major purpose in conducting this study was to determine the effects of response distortion on hiring decisions. Zickar et al. (1996) and Douglas et al. (1996) have both presented simulations showing that response distortion is likely to change the rank ordering of applicants at the upper tail of the distribution of personality scores. These simulations indicate that response distortion should affect who is hired in a top-down hiring system that uses personality inventories as a selection device. This will be particularly true when selection ratios are small (Levin, 1995; Zickar et al., 1996). Adjusting personality scores for the effects of response distortion should change hiring decisions and reduce the number of applicants with high response-distortion scores who are hired.

Hypothesis 4a: Top-down selection of applicants on the basis of personality scores will result in a greater-than-chance proportion of people with high response-distortion scores being selected.
Hypothesis 4b: The rank order of applicants to be hired will change following adjustment of scores to control for response distortion.
Hypothesis 4c: When selection ratios are low, the response-distortion scores of individuals hired using unadjusted personality scores will be significantly higher than those of individuals who would be hired using adjusted scores.

Method

Sample

The sample consisted of 197 job applicants and 73 job incumbents of a property management firm in a Colorado ski resort. The applicant pool was 59% male and predominantly White. Ages of applicants ranged from the early 20s to the mid-50s, with a median in the mid-20s. The distribution of education was bimodal, with a substantial number of applicants having a high school education or less, but also included a sizable number of applicants with college degrees. Demographics for job incumbents were similar to those of job applicants except that only 45% of the job incumbents were male. Job applicants were applying primarily for seasonal positions as laundry attendant, housekeeper, maintenance worker, and front desk clerk. Job incumbents were employed in the same positions.

Measures

Personality. A modified version of the NEO Personality Inventory, Revised (NEO-PI-R; Costa & McCrae, 1992) was used. This is a general-purpose inventory suitable for employment use and a widely recognized measure of the Big Five personality domains. By special permission of the publisher, the NEO-PI-R was modified to include only personality dimensions relevant to this study. (Although this limits our ability to determine the effects of response distortion on all of the Big Five personality dimensions, we believe it is more consistent with good professional practice.) These included the Angry Hostility, Depression, and Impulsiveness facets from the Neuroticism factor; the Warmth, Gregariousness, Excitement-Seeking, and Positive Emotions facets from Extraversion; the Actions facet of the Openness to Experience factor; the Trust, Straightforwardness, Altruism, Compliance, and Tendermindedness facets of the Agreeableness factor, and all facets of the Conscientiousness factor. Respondents were instructed to indicate their level of agreement or disagreement with each statement on a 5-point scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*).

Response distortion. A limitation of many studies of response-distortion bias is the use of social desirability scales that confound intentional distortion with substantive variance related to overall adjustment (Paulhus, 1984, 1991b). The vast majority of studies exploring social desirability have used the Marlowe-Crowne scale (Crowne & Marlowe, 1960), which does not appear to be a pure measure of intentional response distortion (Billings, Guastello, Rieke, & Berkowitz, 1993; Paulhus, 1984; Zerbe & Paulhus, 1987). Similar concerns have been raised about other widely used measures of socially desirable responding, including the Edwards Social Desirability Scale (Edwards, 1957), the Desirability scale from the Personality Research Form (Jackson, 1967), and the K scale from the MMPI (Paulhus, 1991b).

For this study, the Impression Management scale from the Balanced Inventory of Desirable Responding Version 6 (BIDR-IM; Paulhus, 1991a) was used. Despite the name of the scale, this is a relatively "pure" measure of intentional response distortion that has been found to measure conscious faking not related to substantive dimensions of personality that may be

related to the broader construct of impression management (Billings et al., 1993). Additionally, the BIDR-IM appears to have a stable factor structure, with coefficients alpha ranging from .75 to .86 (Paulhus, 1991a). The BIDR-IM items were randomly interspersed with those from the NEO-PI-R, and the same response scale was used for both. The BIDR-IM had an overall mean of 10.4 ($SD = 4.2$) and an estimated internal consistency of $\alpha = .86$.

Procedure

Job applicants completed an inventory containing both the modified NEO-PI-R instrument and the BIDR-IM as one step in a multiple-step selection process that also included a cognitive ability test and an interview with either the personnel director or a department manager. Job incumbents completed the same inventory during work hours, in small groups, over the course of 1 day. Incumbents were told that their scores were being used to create norms for possible future hiring, and they were assured that individual responses would not be made available to management.

Results

Hypothesis 1, stating that job applicants' response-distortion scores would be higher than job incumbents' response-distortion scores, was supported. Response-distortion scores for job applicants ($M = 11.4$, $SD = 4.1$) were significantly higher than those for job incumbents ($M = 7.5$, $SD = 3.0$; $t[269] = 7.6$; $p < .001$). This difference is substantial, representing an effect size of 1.09 standard deviations (Glass, 1977). Another way of describing the size of the difference is to note that 18% of applicants had response-distortion scores that completely exceeded the range of scores received by incumbents. Applicants' scores were very similar to norms reported by Paulhus (1991a) for a "play up good points" (i.e., fake good) condition ($M = 12.3$, $SD = 4.4$), whereas incumbents' scores more closely resembled norms for Paulhus's "respond honestly" condition ($M = 5.8$, $SD = 3.6$).

Hypothesis 2, which dealt with the distribution of response distortion, was also supported. As predicted, there was a negatively skewed distribution (skew $= -.28$) of response-distortion scores among applicants. (The skewness of response-distortion scores among incumbents was .04.) Using the distribution of incumbents as a baseline for relatively honest responses, 29% of job applicants had response-distortion scores two standard deviations above the mean of incumbents and 13% had scores that were three standard deviations above. In fact, two applicants had the maximum score possible on the response-distortion measure. Only 8% of applicants had response-distortion scores one or more sigma below the incumbent mean.

Hypothesis 3a predicted that job applicants' response-distortion scores would be most highly correlated with more apparently job-related personality traits. The data in

Table 1 provide support for this hypothesis. As predicted, Neuroticism (mean $r = .54$) and Conscientiousness (mean $r = .43$) were most strongly correlated with response distortion. Response distortion was also correlated with Agreeableness (mean $r = .38$), although it was less consistently related to Extraversion (mean $r = .08$). As expected, response-distortion scores were not correlated with the single facet of Openness to Experience.

Support for Hypothesis 3b can be found in Table 2, which compares applicants' and incumbents' mean scores on the NEO-PI-R facets. Job applicants had significantly higher means on all facets except one for the positive traits (i.e., Extraversion, Conscientiousness, and Agreeableness) and significantly lower means on all facets for the Neuroticism trait. (The sole exception was for Excitement-Seeking, which in the context of a job application would probably not be viewed as desirable because it may connote risk-taking behavior.) The effect-size column of Table 2 shows that these differences were practically as well as statistically significant, with the average between-group difference equaling .65 standard deviations. Thus Hypothesis 3b was supported, with two exceptions: The differences between applicants and incumbents were somewhat lower than expected for Conscientiousness (mean effect size = .54) and larger than expected for Agreeableness (mean effect size = .73) and Openness to

Table 1
Correlation of Response Distortion With Facets From NEO-PI-R

Facet	Applicants	Incumbents
Extraversion		
Warmth[a]	.25[b]	−.03
Gregariousness	.16[b]	−.04
Excitement Seeking	−.16[b]	−.37[b]
Positive Emotions	.05	−.11
Openness to Experience		
Actions	.08	−.20
Agreeableness		
Trust	.35[b]	.42[b]
Straightforwardness	.50[b]	.48[b]
Altruism	.38[b]	.18
Compliance[a]	.48[b]	.25
Tendermindedness	.21[b]	.12
Conscientiousness		
Competence[a]	.39[b]	.01
Order	.34[b]	.17
Dutifulness	.42[b]	.41[b]
Achievement Striving[a]	.39[b]	.15
Self-Discipline[a]	.49[b]	.25
Deliberation[a]	.57[b]	.39[b]
Neuroticism		
Angry Hostility[a]	−.51[b]	−.27[b]
Depression[a]	−.51[b]	−.18
Impulsiveness	−.60[b]	−.54[b]

Note. NEO-PI-R = NEO Personality Inventory, Revised.
[a] Difference between applicant and incumbent correlations is significant, $p < .05$. [b] Correlation is significantly different from zero, $p < .05$.

Table 2
Applicant and Incumbent Scores on the NEO-PI-R

Facet	Applicants		Incumbents		Effect size
	M	SD	M	SD	
Extraversion					
Warmth	57.6	8.3	50.8	10.0	.74*
Gregariousness	57.1	8.9	51.9	11.3	.51*
Excitement Seeking	56.7	8.6	57.9	10.3	.13
Positive Emotions	56.6	9.1	53.3	10.0	.35*
Openness to Experience					
Actions	57.3	9.0	51.2	9.3	.67*
Agreeableness					
Trust	55.1	8.4	48.3	8.2	.82*
Straightforwardness	53.0	8.7	46.6	9.7	.69*
Altruism	57.1	9.4	49.1	10.3	.81*
Compliance	56.5	9.6	48.0	10.3	.85*
Tendermindedness	54.4	9.4	50.0	8.6	.49*
Conscientiousness					
Competence	56.8	10.2	51.8	9.2	.51*
Order	54.2	9.8	50.9	11.0	.32*
Dutifulness	54.9	10.5	48.6	9.4	.63*
Achievement Striving	57.9	9.6	52.2	10.0	.58*
Self-Discipline	58.3	7.8	51.7	9.3	.77*
Deliberation	53.8	10.7	49.4	8.9	.45*
Neuroticism					
Angry Hostility	47.3	8.7	57.9	9.6	1.16*
Depression	46.0	8.2	53.0	8.6	.83*
Impulsiveness	44.7	9.1	53.2	8.0	.99*

Note. NEO-PI-R = NEO Personality Inventory, Revised. Effect size describes the size of the difference between means in standard deviation units = $|(M_1 + |M_2|)/\sqrt{(SD_1^2 + SD_2^2)/2}|$.
* Difference between applicants and incumbents significant, $p < .05$.

Experience (effect size of .67 for the single facet measured).

These results lead to Hypothesis 4, which predicted that differences in response distortion would affect who is hired. To test this hypothesis, the 162 applicants for whom complete data were available were rank ordered according to their scores on the Conscientiousness factor. (Conscientiousness was chosen because it is the personality dimension most consistently related to job performance and other relevant organizational outcomes (Barrick & Mount, 1991; Ones et al., 1993). Table 3 reports the average Conscientiousness score, the average response-distortion score (in standard score units) among applicants, and the number of applicants with extreme response-distortion scores (defined as scores three or more standard deviations above the incumbent mean) at various selection ratios. With selection ratios of less than approximately 25%, the *average* level of response distortion in applicants hired was at least one standard deviation above the mean; it was nearly two standard deviations above the mean with a selection ratio of 5%. The proportion of applicants with extreme (+3 sigma) levels of response distortion significantly exceeded the base rate of extreme response distortion (13%, or 21 of 162 applicants) at all selection ratios less than 50% (using one-

Table 3

Effect of Selection Ratio (SR) on Conscientiousness and Response Duration (RD)

| SR | No. hired | Conscientiousness | RD (Z score) | High RD | |
				n	%
.05	8	79.6	1.78	7	88
.10	16	76.5	1.22	9	56
.20	32	72.8	1.12	14	44
.30	49	70.2	.80	16	33
.40	65	68.3	.74	17	26
.50	81	66.4	.62	19	23
.60	97	64.6	.58	20	21
.70	113	63.0	.50	21	19
.80	130	61.4	.45	21	16
.90	146	60.1	.34	21	14
1.00	162	56.3	n/a	21	13

Note. Assumes top-down hiring based on Conscientiousness score. Conscientiousness scores are mean *T* scores. RD scores are mean *Z* (standard) RD scores among applicants. High RD = RD scores 3 or more standard deviations above the mean of incumbents.

tailed tests of differences in proportions). In practical terms, if hiring is limited to the top 5% of applicants, 7 of the 8 people hired would have extreme scores on the response-distortion measure. When hiring the top 10%, more than half of the new hires would have similarly skewed scores. These findings confirm Hypothesis 4a.

An important practical question is whether the proportion of applicants hired with high response-distortion scores could be reduced by using response-distortion scores as a control variable. We created a simple test of this possibility by subtracting applicants' response-distortion scores from their Conscientiousness scores. Table 4 presents the mean (uncorrected) Conscientiousness scores, the average response-distortion score (in standard score units) among applicants, and the number of applicants with extreme response-distortion scores who would be hired when applicants were rank-ordered by their adjusted Conscientiousness scores. Comparison of Tables 3 and 4 shows that response distortion had a significant effect on which applicants were hired and that hiring on the basis of the adjusted scores, by way of illustration, would dramatically reduce the average level of response distortion at selection ratios less than approximately 50%.

The rank ordering of job applicants was substantially different when using the adjusted rather than the unadjusted Conscientiousness scores, as predicted in Hypothesis 4b. (The correlation between the adjusted and nonadjusted rank orderings was 50, using Kendall's tau-*b* for ordinal data.) With a selection ratio (SR) of .05, three of the applicants with the highest unadjusted Conscientiousness scores would no longer be hired. All three of these had extreme response-distortion scores (four of the other applicants with extreme response distortion scores would still be hired). On the other hand, the highest ranked

applicant using the adjusted scores was ranked 14th in the original hiring list (because this applicant had a response distortion score .9 *SD* below the mean).

Because applicants with extreme response-distortion scores were particularly overrepresented among the top 5% of applicants in this sample, the effect of using the adjusted Conscientiousness scores is less dramatic when SR = .10. However, the tables show a continuing advantage of the adjusted score when SR = .20. Overall, these data demonstrate that taking response-distortion scores into consideration has a statistically and practically significant effect on who is hired, confirming Hypothesis 4c.

Discussion

These results show that response distortion does occur among real job applicants taking a typical, commercially available test in an actual job selection setting. As we predicted, higher levels of response distortion occurred among job applicants, whereas less response distortion was found among job incumbents in a setting where personality scores would not be made available to their employer. These effects were substantial, with a difference of more than one standard deviation between the mean response-distortion scores of job applicants and job incumbents.

We also found considerable individual differences in response distortion. If all job applicants had responded similarly to the pressure to look good when applying for a job, the effect would simply have been to increase the

Table 4

Effect of Selection Ratio (SR) on Conscientiousness and Response Distortion (RD) When Hiring Is Based on Distortion-Adjusted Conscientiousness Scores

| SR | No. hired | Conscientiousness | RD (Z score) | High RD | | % reduction in high RD |
				n	%	
.05	8	77.4	1.02	4	50	38
.10	16	75.8	.94	8	50	6
.20	32	71.6	.53	8	25	19
.30	49	69.4	.49	11	22	11
.40	65	67.8	.54	15	23	3
.50	81	66.1	.47	17	21	2
.60	97	64.0	.34	17	18	3
.70	113	62.6	.30	18	16	3
.80	130	61.0	.27	19	15	1
.90	146	59.8	.28	21	14	0
1.00	162	56.3	n/a	21	13	

Note. Assumes top-down hiring based on adjusted Conscientiousness score (C − RD). Conscientiousness scores are mean uncorrected *T* scores. RD scores are mean *Z* (standard) scores among applicants. High RD = RD scores 3 or more standard deviations above the mean of incumbents; the last column refers to the decrease in high RD when Conscientiousness scores are adjusted for RD.

average scores for the desirable traits, with no change in the rank ordering of applicants. Whereas this would have important implications for the use of norms in making hiring decisions, it would not greatly affect the correlation between personality scores and criterion measures. Although we did find the expected negative skew in applicants' response-distortion scores, we also found substantial variance, with a considerable number of applicants exhibiting extreme levels of response distortion more than three standard deviations above the mean. It is the variance in response distortion that changes the rank ordering of applicants and potentially lowers the validity and utility of personality scores.

Our results also show that the differences in response distortion translate into differences in personality-scale scores. Applicants had significantly more positive personality profiles than did job incumbents; these differences were robust across personality facets and were substantial in magnitude. The differences were particularly sizable for Neuroticism and Agreeableness, probably due to the socially desirable nature of many of the items in these facets, especially in an employment setting. Of particular importance for employee selection are the sizable differences on the Conscientiousness facets, which are frequently recommended for use in making hiring decisions (Barrick & Mount, 1991; Ones et al., 1993) and which Paulhus et al. (1995) found to be most susceptible to intentional distortion.

These differences between incumbents and applicants on personality-scale scores raise important questions about the construct validity of the scores for employment purposes and for other settings with strong demand characteristics. It seems unlikely that the differences in scores reflect true differences in personality between job applicants and job incumbents. The company had been hiring similar people from the same labor market for years, so it is unlikely that the applicants and incumbents were being sampled from different populations. Moreover, the average tenure of incumbents was short enough to rule out cohort or maturation effects as plausible explanations for the differences. Thus incumbents should have had similar or higher true levels of Conscientiousness and Agreeableness (and perhaps lower levels of Neuroticism) relative to applicants. The more compelling explanation appears to be situational differences in demand characteristics.

The robust correlations between response distortion and personality-scale scores, which in many cases exceeded typical correlations between the personality scales and measures of job performance, raise additional concerns about construct validity. Unadjusted scores on the NEO-PI-R do not appear to provide unbiased measurement of personality, at least for job applicants. Future studies need to address the effects of response distortion

on construct validity, rather than looking only at its effects on criterion-related validity. This is particularly true because of the statistical artifacts that interfere with studying response distortion as a suppressor variable in criterion-related validation studies. By shifting attention instead to the effects of response distortion on individual-level hiring decisions, this study demonstrates that response distortion can have substantial effects on who is hired, particularly when the selection ratio is less than .50. Realizing that with a selection ratio of .05 as many as 88% of the new hires in our sample may have significantly lower true Conscientiousness scores than reported should be a sobering thought for decision makers who thought they were being highly selective.

Future studies need to explore the generalizability of these results to other settings. In our sample, for example, adjustments for response distortion had less effect at a selection ratio of .10 than at selection ratios of .05 or .15, which we speculate was because of the effects of chance with a small sample. We speculate that our results may actually underestimate effects in samples from other settings, with one reason being that the workers included in our sample were likely to be relatively unfamiliar with personality inventories. People applying for jobs in which personality testing is more common (managerial and executive positions, for example) may have a more sophisticated cynicism about such testing and be more willing and able to distort their responses. The nature of the jobs we studied also makes it harder to engage in response distortion, particularly of an extraconventional nature. How much can a person make up about his or her qualifications to be a housekeeper or front-desk clerk?

Implications for Practice

Our results indicate that response distortion needs to be evaluated when using personality testing for employee selection. A more difficult question is how to use information about response distortion when making hiring decisions. We suggest that the appropriate use of this information hinges on the relevance of response distortion to job performance and on future developments in the psychometrics of score adjustments.

A critical issue is whether response distortion reflects job-relevant information or error variance. Some have argued that the ability to distort scores in a socially desirable direction indicates a functional awareness of social norms, which might in fact be related to higher levels of performance (e.g., R. Hogan et al., 1996). If so, our finding that a disproportionate number of the highest ranking applicants were faking their answers is not as serious as we have suggested. We believe there are three aspects of this challenge that need to be considered.

First, Ones et al.'s (1996) meta-analysis casts some

doubt on the position that socially desirable responding is related to job performance. They found that social desirability scores were uncorrelated with measures of task performance ($\rho = .00$), job performance ($\rho = .01$), and counterproductive behavior ($\rho = -.03$), although they were modestly related to ($\rho = .19$) measures of training performance. However, Ones et al. did not consider type of job as a moderator of these relationships.

A reasonable case can be made that the job-relatedness of being able to provide socially desirable responses depends on the nature of the job. For jobs involving a lot of interpersonal interaction, particularly of a somewhat superficial or transitory nature, awareness of social norms and expectations may be functional. One-time interactions with customers, for example, may be facilitated by an ability to provide an appearance of liking the customer and wanting to meet the customer's every need, even when reality is quite the opposite. For other positions, such as an auditor or bank examiner, a predilection for giving the desired rather than the correct answer is not desirable and may be highly dysfunctional. Even for customer service positions, such a tendency may be counterproductive for developing effective long-term relationships that are built on trust. For example, faking on the Personal Characteristics Inventory (Barrick & Mount, 1995) is negatively correlated with job performance in customer service positions (M. Mount, personal communication, September 27, 1996).

Second, the meaning—and functional value—of response distortion depends on the level and nature of the distortion. There are qualitative differences among (a) presenting factually correct information in the most self-aggrandizing fashion, (b) failing to mention negative information, (c) denying negative information, and (d) making up information that will make one look good to achieve a desired end (Levin, 1995). Response distortion is not merely a matter of awareness of social expectations; in its more extreme forms it is also a matter of integrity.

A related issue concerns individual differences in response distortion. Some applicants engaged in much more substantial levels of response distortion than did others, which resulted in their being disproportionately represented among the highest ranked applicants. Most of our analyses focused on individuals whose scores were three or more standard deviations above the mean; some actually received the maximum possible score on the response-distortion measure. To score at that level, these applicants were almost surely engaging in levels of response distortion that far exceeded any reasonable requirement to be socially appropriate or put their best foot forward. Although we have no data to support our conjecture, we speculate that many of the applicants with extreme response-distortion scores would also engage in other forms

of dishonesty. This is a question urgently in need of empirical answers.

At the opposite extreme, test scores and applicant records of applicants exhibiting low levels of response distortion should also be examined carefully. Unless prior analysis has suggested that impression management is relevant to job performance, it is quite possible that the job performance of these individuals will be underpredicted by their personality dimension scores.

One approach is to adjust the scores of all applicants on the basis of their response-distortion scores. This formalizes the process of interpreting scores but presents a dilemma: Is it appropriate to adjust test scores when there may be no evidence that the adjustments increase the validity of the test scores?[1] Although it is not clear that there is any legal mandate to show an increase in validity unless the adjustment itself creates adverse impact (which seems unlikely), this is an issue that deserves more study. Analyses showing the change in who gets hired as a result of response distortion—and that an "honest" person might otherwise lose a job to an applicant who "fakes"— may be quite persuasive to a judge or jury. Studying this issue from the perspective of procedural justice and social validity (Schuler, 1993) might be very enlightening.

Unfortunately, there is little agreement about the optimal statistical approach for adjusting scores (Hough, 1995). Most inventories that include response-distortion measures do little more than recommend caution in interpreting the substantive scores of individuals with high response-distortion scores. Certainly, test scores of applicants exhibiting extreme levels of response distortion should be used with great caution, and it may also be useful to carefully scrutinize all of the materials presented by these applicants for distortions. Given the increased use of personality inventories, the evidence that job applicants do distort scores on personality inventories in a manner that may affect who gets hired, and the likelihood that coaching in how to fake personality tests will become more common (Hough, 1996), there is clearly a need to devote more research attention to the psychometric and social issues underlying more formal response-distortion adjustments. Recent work applying item-response theory to the detection of faking on personality inventories (Zickar & Drasgow, 1996) represents one promising avenue. As researchers and practitioners both become more aware of the threat posed by response distortion to selection systems and to subsequent workforce performance, we are confident that other approaches will be explored as well.

---

[1] We thank an anonymous reviewer for raising this issue.

References

Alliger, G. M., Lilienfield, S. O., & Mitchell, K. E. (1995). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7*, 32–39.

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

Barrick, M. R. & Mount, M. (1995). *The Personal Characteristics Inventory manual.* Unpublished manuscript, University of Iowa, Iowa City.

Barrick, M., & Mount, M. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261–272.

Bass, B. M. (1957). Faking by sales applicants on a forced choice personality inventory. *Journal of Applied Psychology, 41*, 403–404.

Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. *Personnel Psychology, 45*, 389–406.

Billings, S. W., Guastello, S. J., Rieke, M. L., & Berkowitz, M. W. (1993, April). *The content of three measures of socially desirable responding.* Presented at the annual meeting of the Society of Industrial and Organizational Psychology, San Francisco.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, IL: Institute for Personality and Ability Testing.

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47*, 847–860.

Conger, A. J., & Jackson, D. N. (1972). Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement, 32*, 579–599.

Costa, P., & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) professional manual.* Odessa, FL: Psychological Assessment Resources.

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

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

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

Douglas, E. F., McDaniel, M. A. & Snell, A. F. (1996, August). *The validity of non-cognitive measures decays when applicants fake.* Paper presented at the annual meeting of the Academy of Management, Cincinnati, OH.

Drasgow, F., & Kang, T. (1984). Statistical power of differential validity and differential prediction analyses for detecting measurement nonequivalence. *Journal of Applied Psychology, 69*, 498–508.

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

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*, 385–406.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5*, 351–379.

Gough, H. G. (1975). *Manual for the California Psychological Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Hogan, J., & Hogan, R. (1986). *Hogan Personnel Selection Series manual.* Minneapolis, MN: National Computer Systems.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*, 469–477.

Hough, L. (1995, May). *Applicant self descriptions: Evaluating strategies for reducing distortion.* Presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Hough, L. M. (1996a). Can measures of integrity be trusted? *Employment Testing: Law and Policy Reporter*, pp. 97–103, 111.

Hough, L. M. (1996b, April). *Personality measurement and personnel selection: Implementation issues.* Paper presented at the 11th annual conference of the Society for Industrial and Organizational Psychology, San Diego.

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

Hough, L., & Schneider, R. J. (1996). Personality traits, taxonomies, and applications in organizations. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 31–38). San Francisco: Jossey-Bass.

Hunt, S. T., Hansen, T. L., & Paajanen, G. E. (1996). *What do integrity tests measure? An in-depth look at the empirical structure and construct validity of one widely used, predictively valid, personality-based integrity test.* Unpublished manuscript, Ohio State University.

Jackson, D. N. (1967). *Personality Research Form manual.* Goshen, NY: Research Psychologists Press.

Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology, 43*, 48–55.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34–47.

Levin, R. A. (1995, May). *Self-presentation, lies and bullshit: The impact of impression management on employee selection.* Presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Lillibridge, J. R., & Williams, K. J. (1992). Another look at personality and managerial potential: Application of the five-factor model. In K. Kelley (Ed.), *Issues, theory and research in industrial organizational psychology* (pp. 91–115). New York: Elsevier Press.

LoBello, S. G., & Sims, B. N. (1993). Fakability of a commercially produced pre-employment integrity test. *Journal of Business and Psychology, 8*, 265–273.

Mahar, D., Colognon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences, 18*, 605–609.

McCrae, R., & Costa, P. (1983). Social desirability scales:

More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

Ones, D. S., Viswesvaran, C., & Korbin, W. P. (1995, May). *Meta-analyses of fakability estimates: Between-subjects versus within-subjects designs.* Paper presented at the 10th annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

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.

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

Paajanen, G. E. (1987). *The prediction of counterproductive behavior by individual and organizational variables.* Unpublished dissertation, Department of Psychology, University of Minnesota.

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

Paulhus, D. L. (1991a). *Balanced Inventory of Desirable Responding (BIDR) reference manual for version 6.* (Manual available from author at Department of Psychology, University of British Columbia, Vancouver, B. C., Canada V6T 1Y7.)

Paulhus, D. L. (1991b). 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). San Diego: Academic Press.

Paulhus, D. L., & Bruce, M. N. (1991, August). *Faking job profiles.* Paper presented at the annual meeting of the American Psychological Association, San Francisco.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21,* 100–108.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60,* 307–317.

Ryan, A., & Sackett, P. (1987). Pre-employment honesty test-

ing: Fakability, reactions of test takers, and company image. *Journal of Business and Psychology, 1,* 248–256.

Sacheim, H. A., & Gur, R. C. (1978). Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (Vol. 2, pp. 139–197). New York: Plenum Press.

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.

Schuler, H. (1993). Social validity in selection situations: A concept and some empirical results. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 11–26). Hillsdale, NJ: Erlbaum.

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

Villanova, P., & Bernardin, H. J. (1991). Performance appraisal: The means, motive, and opportunity to manage impressions. In R. A. Giacalone & P. Rosenfeld (Eds.), *Applied impression management: How image-making affects managerial decisions* (pp. 81–96). Beverly Hills, CA: Sage.

White, L. A., & Moss, M. C. (1995, May). *Factors influencing concurrent versus predictive validities of personality constructs: Impact of response distortion and item job content.* Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Whyte, W. H., Jr. (1956). How to cheat on personality tests. In *The organization man* (pp. 405–410). New York: Simon & Schuster/Touchstone.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review, 12,* 250–264.

Zickar, M., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement, 20,* 71–87.

Zickar, M., Rosse, J., & Levin, R. (1996, April). *Modeling the effects of faking on personality instruments.* Paper presented at the 11th annual meeting of the Society for Industrial and Organizational Psychology, San Diego.