

Response distortion in personality measurement: born to deceive, yet capable of providing valid self-assessments?

STEPHAN DILCHERT¹, DENIZ S. ONES, CHOCKALINGAM VISWESVARAN & JÜRGEN DELLER

Abstract

This introductory article to the special issue of *Psychology Science* devoted to the subject of “Considering Response Distortion in Personality Measurement for Industrial, Work and Organizational Psychology Research and Practice” presents an overview of the issues of response distortion in personality measurement. It also provides a summary of the other articles published as part of this special issue addressing social desirability, impression management, self-presentation, response distortion, and faking in personality measurement in industrial, work, and organizational settings.

Key words: social desirability scales, response distortion, deception, impression management, faking, noncognitive tests

¹ Correspondence regarding this article may be addressed to Chockalingam Viswesvaran, Department of Psychology, Florida International University, Miami, FL 33199, Email: vish@fiu.edu; to Deniz S. Ones or Stephan Dilchert, Department of Psychology, University of Minnesota, Minneapolis, MN 55455; Emails: deniz.s.ones-1@tc.umn.edu or dilc0002@umn.edu; or Jürgen Deller, University of Lüneburg, Institute of Business Psychology, Wilschenbrucher Weg 84a, 21335 Lüneburg, Germany; Email: deller@uni-lueneburg.de.

Despite some unfounded speculations lacking empirical evidence which have recently surfaced regarding the criterion-related validity of personality scores (Murphy & Dzieweczynski, 2005), personality continues to be a valued predictor in a variety of settings (see Hogan, 2005). Meta-analyses have demonstrated that the criterion-related validity of personality tests is at useful levels for predicting a variety of valued outcomes (Barrick & Mount, 2005; Hough & Ones, 2001; Hurtz & Donovan, 2000; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Dilchert, 2005; Salgado, 1997). Nonetheless, researchers and practitioners have brought about new challenges to personality measurement, especially as part of high-stakes assessments (see Barrick, Mount, & Judge, 2001; Hough & Oswald, 2000). One such challenge is the potential for respondents to distort their responses on non-cognitive measures. For example, job applicants could misrepresent themselves on a personality test hoping to increase the likelihood of obtaining a job offer.

Response distortion is certainly not confined to its effects on the criterion-related validity of self-reported personality test scores (Moorman & Podsakoff, 1992; Randall & Fernandes, 1991). In general, response distortion seems to be a concern in any assessment conducted for the purposes of distributing valued outcomes (e.g., jobs, promotions, development opportunities), regardless of the testing tool or medium (e.g., personality test, interview, assessment center). In fact, all high-stakes assessments are likely to elicit deception from assesseees. Understanding the antecedents and consequences of socially desirable responding in psychological assessments has scientific and practical value for deciphering what data really mean, and how they are best used operationally.

The primary objective of this special issue was to bring together an international group of researchers to provide empirical studies and thoughtful commentaries on issues surrounding response distortion. The special issue aims to address social desirability, impression management, self-presentation, response distortion, and faking in non-cognitive measures, especially personality scales, in occupational settings. We begin this introduction by sketching the importance of the topic to the science and practice of industrial, work, and organizational (IWO) psychology. Following this, we highlight a few key issues on socially desirable responding. We conclude with a brief summary of the different articles included in this special issue.

Relevance of response distortion to IWO psychology

Throughout the last century, IWO psychology researchers and practitioners have expended much effort to identify individuals who would be “better” performers on the job (Guion, 1998). Better performance has been defined in a variety of ways, ranging from a narrow focus on task performance to more inclusive approaches including citizenship behaviors, the absence of counterproductive behaviors, and satisfaction of employees with the quality of work life (see Viswesvaran & Ones, 2000; Viswesvaran, Schmidt, & Ones, 2005). In this endeavor, individual differences in personality traits have been found to correlate with performance in general, as well as specific aspects of performance. Evidence of criterion-related validity has encouraged IWO psychologists and organizations to assess personality traits of applicants and hire those with the traits known to correlate with (and predict) later performance on the job. When assessments are made for the purpose of making decisions that affect the employment of test takers (e.g., getting a job or not, getting promoted or not,

et cetera), those assessed have an incentive to produce a score that will increase the likelihood of the desired outcome. Presumably, this is the case even if that score would not reflect their true nature (i.e., would require distortion of the truth or outright lies). The discrepancy between respondents' true nature and observed predictor scores in motivating settings is commonly thought to be due to intentional response distortion or "faking". Distorted responses are assumed to stem from a desire to manage impressions and present oneself in a socially desirable way.

It has been suggested that response distortion could destroy or severely attenuate the utility of non-cognitive assessments in predicting the performance of employees (i.e., lower criterion-related validity of scores). If personality assessments are used such that the respondents are rank-ordered based on their scores, and if this rank-ordering is used to make decisions, individual differences in socially desirable responding could distort rank order and consequently influence high-stakes decisions. Hence, whether individual differences in socially desirable responding exist among applicants, and whether such differences affect the criterion-related validity of personality scales, becomes a key concern. Several studies (e.g., Hough, 1998) have addressed this issue directly, and a meta-analytic cumulation of such studies (Ones, Viswesvaran, & Reiss, 1996) has found that socially desirable responding, as assessed by social desirability scales, does not attenuate the criterion-related validity of personality test scores. Similarly, the hypothesis that social desirability moderates personality scale validities has not received any substantial support (cf. Hough, Eaton, Dunnette, Kamp, & McCloy, 1990). Personality scales, even when used for high-stakes assessments, retain their criterion-related validity (see, for example, Ones, Viswesvaran, & Schmidt, 1993, for criterion-related validities of integrity tests for job applicant samples).

It has been suggested that socially desirable responses could also function as predictors of job performance in that those individuals high on social desirability would also excel on the job, especially in occupations requiring interpersonal skills (Marcus, 2003; Nicholson & Hogan, 1990). However, the limited empirical evidence to date (e.g., Jackson, Ones, Sianangil, this issue, Viswesvaran, Ones, & Hough, 2001) has found little support for social desirability scales predicting job performance or its facets, even for jobs and situations that would benefit from impression management.

Yet, the issue of socially desirable response distortion goes beyond potential attenuating effects on the criterion-related validity of personality scale scores derived from self-reports. All communication involves deception to a certain degree (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). Hogan (2005) argues that the proclivity to deceive is an evolutionary mechanism that helps humans gain access to life's resources. Understanding socially desirable responding and response distortion in all communications thus requires going beyond its influence on criterion-related validity. Socially desirable responding is an inherent part of human nature, and viewed in that context, socially desirable responding is an important topic for our science. In the next section, we discuss some of the issues that have been researched in this area.

Research on faking, response distortion, and social desirability

Paulhus (1984) made the distinction between impression management (which involves an intention to deceive others) and self-deception. Self-deception is defined as “any positively biased response that respondents actually believe to be true”. While this distinction has conceptual merit, it has not always received empirical support. Substantial correlations have been reported between the two factors of social desirable responding (cf. Ones et al., 1996).

In addressing the role of response distortion in self-report personality assessment, four questions have been addressed in the literature. First, researchers have investigated whether respondents can distort their responses and if so, whether there are individual differences in the ability and motivation to distort (e.g., McFarland & Ryan, 2000). Second, researchers have attempted to answer the question whether respondents do in fact distort their responses in actual high-stakes settings (e.g., Donovan, Dwight, & Hurtz, 2003). Third, researchers have investigated the effects of such distortion on the criterion-related validity and construct validity of the resulting scores (e.g., Ellingson, Smith, & Sackett, 2001; Stark, Chernyshenko, Chan, Lee, & Drasgow, 2001). Finally, researchers have proposed palliatives and tested their effectiveness to control socially desirable responding (e.g., Holden, Wood, & Tomashewski, 2001; Hough, 1998, see below).

Can respondents fake?

To address the question of whether individuals can fake their responses in a socially desirable way, lab studies in which responses are directed (e.g., to fake good, fake bad, answer like a job applicant) are most commonly utilized, as they are presumed to reveal the upper limits of response distortion. In lab studies of directed faking, researchers have employed either a between subjects or within subjects design (Viswesvaran & Ones, 1999). In a between subjects design, one group of participants is asked to respond in an honest manner, whereas another group responds under instructions to fake or to portray themselves as desirable job applicants. In a within-subjects design the same individuals take a given test twice, once under instructions to distort and once under instructions to produce honest responses. Of these two approaches, the within subjects design is widely regarded to be the superior design as it controls for individual differences correlates of faking. However, the within-subjects design is also susceptible to testing effects (cf. Hausknecht, Trevor, & Farr, 2002). Meta-analytic cumulation of this literature provides clear support for the conclusion that respondents can distort their responses if they are instructed to do so (Viswesvaran & Ones, 1999). Furthermore, research has also shown that there are individual differences in the ability and motivation to fake (McFarland & Ryan, 2000).

Do applicants fake?

The more intractable question is whether respondents in high-stakes assessments actually provide socially desirable responses. To address this question, some researchers have compared scores of applicants to job incumbents. Data showing that applicants score higher on non-cognitive predictor scales have been touted as evidence that respondents in high-stakes

testing distort their responses. Yet, the attraction-selection-attrition (ASA) framework provides a theoretical rationale for applicant-incumbent differences that does not rely on the response distortion explanation (Schneider, 1987). Comparisons of applicants and incumbents also ignore the finding that the personality of individuals may change with socialization and experiences on the job (see Schneider, Smith, Taylor, & Fleenor, 1998).

There are also psychometric concerns around applicant-incumbent comparisons. Direct and indirect range restriction influences the scores of incumbents to a greater extent than those of applicants. Also, there might be measurement error differences between incumbent and applicant samples: It is quite plausible that employees are not as motivated as applicants to avoid careless responding, which could lead to increased unreliability of scores in incumbent samples. These issues further complicate the use of applicant-incumbent group mean-score differences as indicators of situationally determined faking behavior. While the apparent group differences between applicants and incumbents remain a concern for many non-cognitive measures, especially with regard to the use of appropriate norms (Ones & Viswesvaran, 2003), actual differences might be less drastic than commonly assumed. Ones and Viswesvaran (2001) documented that even extreme selection ratios and labor market conditions do not result in inordinately large group mean-score differences between applicants and incumbents.

Another way of addressing the question at hand is to compare personality scores obtained during the application process with scores obtained when a test is re-administered (Griffeth, Rutkowski, Gujar, Yoshita, & Steelman, 2005). However, this approach is hampered by error introduced due to potential differences in other response sets (e.g., carelessness). Nonetheless, strategies that make use of assessments obtained for different purposes might provide a valuable insight into the issue of situationally determined response distortion. For example, Ellingson, Sackett, and Connelly (in press) examined personality scale score differences between individuals completing a personality test both for selection and for developmental purposes; few differences were found.

When evaluating the question of whether individuals fake, the construct validity of measures used to assess the degree of response distortion (e.g., social desirability scales) also needs to be considered. There is now strong evidence that scores on social desirability scales carry substantive trait variance, and thus reflect stable individual differences in addition to situation specific response behaviors. Personality correlates of socially desirable response tendencies are emotional stability, conscientiousness, and agreeableness (Ones et al., 1996). Cognitive ability is also a likely (negative) correlate of socially desirable responding, particularly in job application settings where candidates have the motivation to use their cognitive capacity to subtly provide the most desirable responses on items measuring key constructs (Dilchert & Ones, 2005).

Anecdotal evidence from the users of tests and perceptions of applicants has been presented as evidence that applicants do fake their responses in a socially desirable way (Rees & Metcalfe, 2003). While interesting, we caution against using such unsystematic and subjective reports as evidence that applicants do indeed fake on commonly used personnel selection measures. As long as such subjective impressions are not backed up by empirical evidence, they will remain pseudo-scientific lore.

Unfortunately, to present evidence showing that applicants do not fake in a given situation is just as intricate a challenge as to present evidence for the opposite. We are not arguing that the absence of evidence is equivalent to evidence of absence; we are merely pointing

out that this question has yet to be satisfactorily answered by our science. Given the prevalent belief among many organizational stakeholders that applicants can and do fake their responses, the issue of response distortion needs to be adequately addressed if we want to see state-of-the-art psychological tools be utilized and make a difference in real-world settings.

Further, given the pervasive influence of socially desirable responding on all types of communications in everyday life, it is certainly beneficial to examine process models of how responses are generated. Hogan (2005) argues that socially desirable responding can be conceptualized as attempts by the respondents to negotiate an identity for themselves. This socio-analytic view of responding to test items suggests that one should not worry about “truthful” responses but rather focus on the validity of the responses for the decisions to be made (Hogan & Shelton, 1998). McFarland and Ryan (McFarland & Ryan, 2000) also discuss a model of faking that takes into account trait and situational factors that influence socially desirable responding. Future work in this area would benefit from considering whether such process models of faking apply beyond personality self report assessments (i.e., to scores obtained in interviews, assessment center exercises, and so forth).

What are the effects of response distortion?

The third question that has been investigated is the effect of socially desirable responding on the criterion-related and construct validity of personality scale scores. Personality measures display useful levels of predictive validity for job applicant samples (Barrick & Mount, 2005). Despite potential threats of response distortion, the criterion-related validity of personality measures in general remains useful for personnel selection and even substantial for most compound personality scales such as integrity tests, customer service scales, and managerial potential scales (Ones et al., 2005).

It has been argued that the correlation coefficient is not a sensitive indicator of rank order changes at the extreme ends of the trait distribution and thus does not pick up the effects of faking on who gets hired (Rosse, Stecher, Miller, & Levin, 1998). Drasgow and Kang (1984) found that correlation coefficients are not affected by rank ordering changes in the extreme ends of one of the two bivariate distributions of variables being correlated. Extending this argument, some researchers suggested that while criterion-related validity may not be affected, the list of individuals hired in personnel selection settings will be different such that fakers are more likely to be selected (Mueller-Hanson, Heggstad, & Thornton, 2003; Rosse et al., 1998). Although plausible, empirical research on this hypothesis has been confined to experimental situations; in fact, the proposed hypothesis may very well be untestable in the field. Thus, we are left with strong empirical evidence for substantial validities in high-stakes assessments and the untestable hypothesis that some of those selected might have scored lower (and thus potentially not obtained a job) had they not engaged in response distortion.

The effects of social desirability on the construct validity of personality measures have also been addressed. Multiple group analyses (with groups differing on socially desirable responding) have been conducted. Groups have been defined based on situational (motivating versus non-motivating settings) or social desirability trait levels (based on impression management scales). Analytically, different approaches such as confirmatory factor analysis and item response theory have been employed. Though not unequivocal, the preponderance

of evidence from this literature points to the following conclusions: Socially desirable responding does not deteriorate the factor structure of personality measures (Ellingson et al., 2001). When IRT approaches are used to investigate the same question, differential item functioning for motivated versus non-motivated individuals is found occasionally (Stark et al., 2001), yet there is little evidence of differential test functioning (Henry & Raju, this issue, Stark et al., 2001). In general, then, socially desirable responding does not appear to affect the construct validity of self-reported personality measures (see also Bradley & Hauenstein, this issue).

Although socially desirable responding does not affect the construct validity or criterion-related validity of personality assessments, there is still a uniformly negative perception about the prevalence and effects of socially desirable responding (Alliger, Lilienfeld, & Mitchell, 1996; Donovan et al., 2003). More research is needed to address the antecedents and consequences of such perceptions held by central stakeholders in psychological testing. Fairness perceptions are important, and we need to consider the reactions that our measures elicit among test users as well as the lay public. In addition, we need to examine how organizational decision-makers, lawyers, courts (and juries) react to fakable psychological tools, and suggested palliatives, decision-making heuristics, et cetera.

Approaches to dealing with social desirability bias

Several strategies have been proposed to deal with the effects of socially desirable responding. At the outset we should acknowledge that most of these strategies, investigated individually, do not seem to have lasting substantial effects. Whether a combination of these different practices will be more effective requires further investigation (see Ramsay, Schmitt, Oswald, Kim, & Gillespie, this issue).

Approaches to dealing with social desirability bias can be classified by their purpose: Strategies have been proposed to *discourage* applicants from engaging in response distortion. Additionally, *hurdle* approaches have been developed to make it more difficult to distort answers on psychological measures. Finally, means have been proposed to *detect* response distortion among test takers.

Approaches to discourage distortion. Honesty warnings are a straightforward means to discourage response distortion and typically take one of two forms: Warnings that faking on a measure can be identified or that faking will have negative consequences. A recent quantitative summary showed that warnings have a relatively small effect in reducing mean-scores on personality measures ($d = .23$, $k = 10$) (Dwight & Donovan, 2003). An important question is whether such warnings provide a better approximation of the true score or introduce biases of their own. Warnings might generate negative applicant reactions; future research is needed to gather the data needed to address this issue. In one of the few studies on this topic, McFarland (2003) found that warnings did not have a negative effect on test-taker perceptions. However, more research on high-stakes situations is needed. It has been suggested that warnings also have the potential to negatively impact the construct validity of personality measures (Vasilopoulos, Cucina, & McElreath, 2005). We want to acknowledge that to the extent that warning-type instructions represent efforts to ensure proper test taking and administration, such instructions should be adopted. For example, encouraging respondents to

provide truthful responses, as well as efforts to minimize response sets such as acquiescence or carelessness, are likely to improve the testing enterprise.

Hurdle approaches. Much effort has been spent on devising hurdles to socially desirable responding. One approach has been to devise subtle assessment tools using empirical keying, both method specific (e.g., biodata, see Mael & Hirsch, 1993) as well as construct based (e.g., subtle personality scales, see Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001). Some authors have suggested that in high-stakes settings, we should generally attempt to assess personality constructs which are poorly understood by test takers and thus more difficult to fake (Furnham, 1986). Even if socially desirable responding is not a characteristic of test items (but a tendency of the respondent), such efforts may reduce the ability of respondents to distort their responses. Empirical research (cf. Viswesvaran & Ones, 1999) does not support the hypothesis that covert or subtle items are less susceptible to socially desirable responding. Also, subtle personality scales are not always guaranteed to yield better validity (see Holden & Jackson, 1981, for example). Other approaches might prove to be more fruitful: Recently, McFarland, Ryan, and Ellis (2002) showed that scores on certain personality scales (neuroticism and conscientiousness) are less easily faked when items are not grouped together.

Asking test takers to elaborate on their responses might also be a viable approach to reduce response distortion on certain assessment tools. Schmitt et al. (2003) have shown that elaboration on answers to biodata questions reduces mean scores. However, no notable changes in correlations of scores with social desirability scale scores or criterion scores were observed.

Another hurdle approach has been to devise test modes and item response formats that increase the difficulty of faking on a variety of measures. The most promising strategy entails presenting response options that are matched for social desirability using predetermined endorsement frequencies or desirability ratings (cf. Waters, 1965). This technique is commonly employed in forced-choice measures that ask test takers to choose among a set of equally desirable options loading on different scales/traits (e.g., Rust, 1999). However, this approach has been heavily criticized, as it yields fully or at least partially ipsative scores (cf. Hicks, 1970), which pose psychometric challenges as well as threats to construct validity (see Dunlap & Cornwell, 1994; Johnson, Wood, & Blinkhorn, 1988; Meade, 2004; Tenopyr, 1988).

Other approaches have centered on the test medium as a means of reducing social desirability bias. For example, it has been suggested that altering the administration mode of interviews (from contexts requiring interpersonal interaction to non-social ones) would reduce social desirability bias (Martin & Nagao, 1989). Richman, Kiesler, Weisband, and Drasgow (1999) meta-analytically reviewed the effect of paper-and-pencil, face-to-face, and computer administration of non-cognitive tests as well as interviews. Their results suggest that there are few differences between administration modes with regard to mean-scores on the various measures.

Detecting response distortion. Response time latencies have been used to identify potentially invalid test protocols and detect response distortion. While some of these attempts have been quite successful at distinguishing between individuals in honest and fake good conditions in laboratory studies (Holden, Kroner, Fekken, & Popham, 1992), the efficiency of this approach in personnel assessment settings remains to be demonstrated. Additionally, there is evidence that response latencies cannot be used to detect fakers on all assessment tools

(Kluger, Reilly, & Russell, 1991) and that their ability to detect fakers is susceptible to coaching (Robie et al., 2000).

Specific scales have been constructed that are purported to assess various patterns of response distortion (e.g., social desirability, impression management, self deception, defensiveness, faking good). Many practitioners continue to rely on scores on these scales in making a determination about the degree of response distortion engaged in by applicants (Swanson & Ones, 2002). With regard to their construct validity, it should be noted that such scale scores correlate substantially with conscientiousness and emotional stability (see above). The fact that these patterns are also observed in non-applicant samples, as well as in peer reports of personality has led many to conclude that these scales in fact carry “more substance than style” (McCrae & Costa, 1983).

Potential palliatives

There are only few available remedies to the problem of social desirability bias. One obvious approach is the identification of “fakers” by one of the means reviewed above, and the subsequent readministration of the selection tool after more severe warnings have been issued. Another, more drastic approach is the exclusion of individuals from the application process who score above a certain (arbitrary) threshold on social desirability scales. Yet another strategy is the correction of predictor scale scores based on individuals’ scores on such scales.

We believe that using social desirability scales to identify fakers or to correct personality scale scores is not a viable solution to the problem of social desirability bias. In evaluating this approach, several questions need to be asked. If test scores are corrected using scores on a social desirability scale, do these corrected scores approximate individuals’ responses under “honest” conditions? Do corrections for social desirability improve criterion-related validities? Do social desirability scales reflect only situation specific response distortion, or do they also reflect true trait variance? Does correcting scores for social desirability have negative implications?

It has previously been shown that corrected personality scores do not approximate individuals’ honest responses. Ellingson, Sackett, and Hough (1999) used a within-subjects design to compare honest scores to those obtained under motivating conditions (“respond as an applicant”). While correcting scores for social desirability reduced group mean-scores, the effect on individuals’ scores was unsystematic. The proportion of individuals correctly identified as scoring high on a trait did not increase when scores were corrected for social desirability in the applicant condition. These findings have direct implications for personnel selection, as they make it unlikely that social desirability corrections will improve the criterion-related validity of personality scales. Indeed, this has been confirmed empirically (Christiansen, Goffin, Johnston, & Rothstein, 1994; Hough, 1998; Ones et al., 1996).

The potential harm caused by social desirability corrections also needs to be addressed. Given that social desirability scales reflect true trait variance, partialling out social desirability from other personality scale scores might pose a threat to the construct validity of adjusted scores. Additionally, the potential of social desirability scales to create adverse impact has been raised and empirically linked to the negative cognitive ability-social desirability relationship in personnel selection settings (Dilchert & Ones, 2005).

Forced-choice items have been advocated as a means to address socially desirable responding for decades (see Bass, 1957). However, forced-choice scales introduce ipsativity, raising concerns regarding inter-individual comparisons required for personnel decision making. Tenopyr (1988) showed that scale interdependencies in ipsative measures result in major difficulties in estimating reliabilities. The relationship between ipsative and normative measures of personality is often weak, resulting in different selection decisions depending on what type of measure is used for personnel selection (see Meade, 2004, for example). However, in some investigations, forced-choice scales displayed less score inflation than normative scales under instructions to fake (Bowen, Martin, & Hunt, 2002; Christiansen, Burns, & Montgomery, 2005). Some empirical studies (e.g., Villanova, Bernardin, Johnson, & Dahmus, 1994) have occasionally reported higher validities for forced-choice measures but it is not clear (1) whether the predictors assess the intended constructs or (2) whether the incremental validity is due to the assessment of cognitive ability which potentially influences performance on forced-choice items.

The devastating effects of ipsative item formats on reliability and factor structure of non-cognitive inventories continues to be a major concern with traditionally scored forced-choice inventories (see Closs, 1996; Johnson et al., 1988). However, IRT based techniques have been developed to obtain scores with normative properties from such items. At least two such approaches have been presented recently (McCloy, Heggstad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & Williams, 2006). Tests of both methods have confirmed that they can be used to estimate normative scores from ipsative personality scales (Chernyshenko et al., 2006; Heggstad, Morrison, Reeve, & McCloy, 2006). Yet, the utility of such methods for personnel selection is still unclear. For example, while the multidimensional forced-choice format seems resistant to score inflation at the group level, the same is not true at the individual level (McCloy et al., 2005). Future research will have to determine whether such new and innovative formats can be further improved in order to prevent individuals from distorting their responses without compromising the psychometric properties or predictive validity of proven personnel selection tools.

Overview of papers included in this issue

Marcus (this issue) addresses apparent contradictions expressed in the literature on the effect of socially desirable responding on selection decisions. As noted earlier, correlation coefficients expressing criterion-related validity are not affected by socially desirable responding. Nonetheless, some researchers have claimed that the list of individuals hired using a given test may change as a result of deceptive responses (see above). Marcus (this issue) examines the parameters that influence rank ordering at the top end of the score distribution, as well as validity coefficients and concludes that both rank order and criterion-related validity are similarly sensitive to variability in faking behavior.

What is social desirability? One perspective is that social desirability is whatever social desirability scales measure. Others before us have stressed the importance of establishing the nomological net of measures of social desirability (see Nicholson & Hogan, 1990). In the next paper of this special issue, Mesmer-Magnus, Viswesvaran, Deshpande, and Jacob investigate the correlates of a well-known measure of social desirability – the Marlowe-Crowne scale – with measures of self-esteem, over-claiming, and emotional intelligence. Emotional

intelligence is a concept that has gained popularity in recent years (Goleman, 1995; Mayer, Salovey, & Caruso, 2000); its construct definition incorporates several facets of understanding, managing, perceiving, and reasoning with emotions to effectively deal with a given situation (Law, Wong, & Song, 2004; Mayer et al., 2000). The authors report a correlation of .44 between emotional intelligence and social desirability scale scores. Social desirability has been defined as an overly positive presentation of the self (Paulhus, 1984); however, the correlation with the over-claiming scale in Mesmer-Magnus et al.'s sample was only $-.05$.

Jackson, Ones and Sinangil (this issue) demonstrate that social desirability scales do not correlate well with job performance or its facets among expatriates, adding another piece to our knowledge of the nomological net of social desirability scales. Social desirability scales are neither predictive of job performance in general (Ones et al., 1996) nor performance facets among managers (Viswesvaran et al., 2001). It is reasonable to inquire whether such scales predict adjustment or performance facets for expatriates (Montagliani & Giacalone, 1998). Jackson, Ones and Sinangil investigate whether individual differences in social desirability are related to expatriate adjustment. They also test whether there are mediating roles for adjustment in the effect of social desirability on actual job performance. A strength of this study is the non self-report criteria utilized to assess performance.

Henry and Raju (this issue) investigate the effects of socially desirable responding on the construct validity of personality scores. As noted earlier in this introduction, Ellingson et al. (2001) found no deleterious effects of impression management on construct validity whereas Schmit and Ryan (1993) found that construct validity was affected. Stark et al. (2001) have advanced the hypothesis that conflicting findings are due to the use of factor analyses and have suggested that an IRT framework is required. Henry and Raju (this issue), using a large sample and an IRT framework to investigate this issue, find that concerns regarding the possibility of large-scale measurement inequivalence between those scoring high and low on impression management scales are not supported.

Researchers and practitioners have proposed various palliatives to address the issue of response distortion (see above). Ramsay et al. (this issue) investigate the additive and interactive effects of motivation, coaching, and item formats on response distortion on biodata and situational judgment test items. They also investigate the effect of requiring elaborations of responses. Such research assessing the cumulative effects of suggested palliatives represents a fruitful avenue for future work in this area.

Mueller-Hanson, Heggstad, and Thornton (this issue) continue the exploration of the correlates of directed faking behavior by offering an integrative model. They find that personality traits such as emotional stability and conscientiousness as well as perceptions of situational factors correlate with directed faking behavior. Interestingly, the authors find that some traits such as conscientiousness are related to both willingness to distort responses and perceptions of the situation (i.e., perceived importance of faking, perceived behavioral control, and subjective norms). This dual role of stable personality traits reinforces the hypothesis that individuals are not passive observers but actively create their own situation. If so, it is doubly difficult to disentangle the effects of the situation and individual traits on response distortion individuals engage in.

Bradley and Hauenstein (this issue) attempt to address previously conflicting findings (Ellingson et al. 2001; Schmit & Ryan, 1993; Stark et al. 2001) by separately cumulating (using meta-analysis) the correlations between personality dimensions obtained in applicant and incumbent samples. The authors report very minor differences in factor loadings by

sample type, suggesting that construct validity of personality measures is robust to assumed response distortion tendencies. This is an important study in this field and should assuage concerns of personality researchers and practitioners about using personality scale scores in decision-making.

Koenig, Melchers, Kleinmann, Richter, Vaccines, and Klehe (this issue) investigate a novel idea that the ability to identify evaluative criteria is related to test scores used in personnel decision-making. They test their hypothesis in the domain of integrity tests, a domain in which response distortion is relatively rarely investigated. A feature of their study is the use of a German sample which – in combination with the Jackson et al. and Khorramdel and Kubinger articles in this issue – presents rare data from samples outside the North American context. Given the potential influences of culture on self-report measures (e.g., modesty bias), more data from different cultural regions would be a welcome addition to the literature on socially desirable responding.

Finally, Khorramdel and Kubinger (this issue) examine the influence of three variables on personality scale scores: time limits for responding, response format (dichotomous versus analogue), and honesty warnings. This is a useful demonstration that multiple features influence responses to personality items and that studying socially desirable responding in isolation is an insufficient approach to modelling reality in any assessment context.

Taken together, the nine articles in this special issue cover the gamut of research on response distortion and socially desirable responding on measures used to assess personality. The articles address definitional issues, the nomological net of social desirability measures, factors influencing response distortion (both individual traits and situational factors), and the effects of response distortion on criterion and construct validity of personality scores. Rich data from different continents and from large samples are presented along with different analytical techniques (SEM, meta-analysis, regression, IRT) and investigations using different assessment tools. We hope this special issue provides a panoramic view of the field and spurs more research to help industrial, work, and organizational psychologists better understand self reports in high-stakes settings.

References

- Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science*, 7, 32-39.
- Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, 18, 359-372.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-30.
- Bass, B. M. (1957). Faking by sales applicants of a forced choice personality inventory. *Journal of Applied Psychology*, 41, 403-404.
- Bowen, C.-C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis*, 10, 240-259.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2). Manual for administration, scoring, and interpretation (Revised ed.)*. Minneapolis, MN: University of Minnesota Press.

- Chernyshenko, O. S., Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R. B., & Tuttle, M. D. (2006, May). Normative score comparisons from single stimulus, unidimensional forced choice, and multidimensional forced choice personality measures using item response theory. In O. S. Chernyshenko (Chair), *Innovative response formats in personality assessments: Psychometric and validity investigations*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, Dallas, Texas.
- Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, 18, 267-307.
- 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.
- Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational & Organizational Psychology*, 69, 41-47.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70, 979-995.
- Dilchert, S., & Ones, D. S. (2005, April). Another trouble with social desirability scales: g-fueled race differences. Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, California.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, 16, 81-106.
- Dragow, F., & Kang, T. (1984). Statistical power of differential validity and differential prediction analyses for detecting measurement nonequivalence. *Journal of Applied Psychology*, 69, 498-508.
- Dunlap, W. P., & Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, 29, 115-126.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1-23.
- Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (in press). Do applicants distort their responses? Personality assessment across selection and development contexts. *Journal of Applied Psychology*.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, 84, 155-166.
- Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, 86, 122-133.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7, 385-400.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam.
- Griffeth, R. W., Rutkowski, K., Gujar, A., Yoshita, Y., & Steelman, L. (2005). Modeling applicant faking: New methods to examine an old problem. Manuscript submitted for publication.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hausknecht, J. P., Trevor, C. O., & Farr, J. L. (2002). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology*, 87, 243-254.

- Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91*, 9-24.
- Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*, 167-184.
- Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance, 18*, 331-341.
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance, 11*, 129-144.
- Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology, 37*, 379-386.
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology, 63*, 272-279.
- Holden, R. R., Wood, L. L., & Tomaszewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology, 81*, 160-169.
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11*, 209-244.
- 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., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H.K. Sinangil & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology (Vol. 1: Personnel psychology, pp. 233-277)*. London, UK: Sage.
- Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future--Remembering the past. *Annual Review of Psychology, 51*, 631-664.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869-879.
- Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouiser and spuriouiser: The use of ipsative personality tests. *Journal of Occupational Psychology, 61*, 153-162.
- Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? *Journal of Applied Psychology, 76*, 889-896.
- Law, K. S., Wong, C.-S., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology, 89*, 483-496.
- Mael, F. A., & Hirsch, A. C. (1993). Rainforest empiricism and quasi-rationality: Two approaches to objective biodata. *Personnel Psychology, 46*, 719-738.
- Marcus, B. (2003). Persönlichkeitstests in der Personalauswahl: Sind "sozial erwünschte" Antworten wirklich nicht wünschenswert? [Personality testing in personnel selection: Is "socially desirable" responding really undesirable?]. *Zeitschrift für Psychologie, 211*, 138-148.
- Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology, 74*, 72-80.
- Mayer, J. D., Salovey, P., & Caruso, D. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence (pp. 396-420)*. New York: Cambridge University Press.

- McCloy, R. A., Heggstad, E. D., & Reeve, C. L. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, 8, 222-248.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51, 882-888.
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment*, 11, 265-276.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, 85, 812-821.
- McFarland, L. A., Ryan, A. M., & Ellis, A. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment*, 78, 348-369.
- Meade, A. W. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, 531-552.
- Montagliani, A., & Giacalone, R. A. (1998). Impression management and cross-cultural adaptation. *Journal of Social Psychology*, 138, 598-608.
- 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 behaviour research. *Journal of Occupational and Organizational Psychology*, 65, 131-149.
- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology*, 88, 348-355.
- Murphy, K. R., & Dziewieczynski, J. L. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, 18, 343-357.
- Nicholson, R. A., & Hogan, R. (1990). The construct validity of social desirability. *American Psychologist*, 45, 290-292.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11, 245-269.
- Ones, D. S., & Viswesvaran, C. (2001, August). Job applicant response distortion on personality scale scores: Labor market influences. Poster session presented at the annual conference of the American Psychological Association, San Francisco, CA.
- Ones, D. S., & Viswesvaran, C. (2003). Job-specific applicant pools and national norms for personality scales: Implications for range-restriction corrections in validation research. *Journal of Applied Psychology*, 88, 570-577.
- Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, 18, 389-404.
- 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. 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. *Journal of Applied Psychology*, 78, 679-703.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.

- Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, 10, 805-817.
- Rees, C. J., & Metcalfe, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology*, 18, 156-165.
- Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84, 754-775.
- Robie, C., Curtin, P. J., Foster, T. C., Phillips, H. L., Zbylut, M., & Tetrick, L. E. (2000). The effects of coaching on the utility of response latencies in detecting fakers on a personality measure. *Canadian Journal of Behavioural Science*, 32, 226-233.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644.
- Rust, J. (1999). The validity of the Giotto integrity test. *Personality and Individual Differences*, 27, 755-768.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 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.
- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., Ramsay, L. J., & Yoo, T.-Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology*, 88, 979-988.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437-453.
- Schneider, B., Smith, D. B., Taylor, S., & Fleener, J. (1998). Personality and organizations: A test of the homogeneity of personality hypothesis. *Journal of Applied Psychology*, 83, 462-470.
- Stark, S., Chernyshenko, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86, 943-953.
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model. *Applied Psychological Measurement*, 29, 184-203.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 91, 25-39.
- Swanson, A., & Ones, D. S. (2002, April). The effects of individual differences and situational variables on faking. Poster session presented at the annual conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.
- Tenopyr, M. L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology*, 73, 749-751.
- Vasilopoulos, N. L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *Journal of Applied Psychology*, 90, 306-322.
- Villanova, P., Bernardin, H. J., Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnel. *Personnel Psychology*, 47, 73-90.

- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197-210.
- Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, 8, 216-226.
- Viswesvaran, C., Ones, D. S., & Hough, L. M. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment*, 9, 277-289.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, 90, 108-131.
- Waters, L. K. (1965). A note on the "fakability" of forced-choice scales. *Personnel Psychology*, 18, 187-191.