

IN SUPPORT OF PERSONALITY ASSESSMENT IN ORGANIZATIONAL SETTINGS

DENIZ S. ONES AND STEPHAN DILCHERT

Department of Psychology
University of Minnesota

CHOCKALINGAM VISWESVARAN

Department of Psychology
Florida International University

TIMOTHY A. JUDGE

Department of Management
University of Florida

Personality constructs have been demonstrated to be useful for explaining and predicting attitudes, behaviors, performance, and outcomes in organizational settings. Many professionally developed measures of personality constructs display useful levels of criterion-related validity for job performance and its facets. In this response to Morgeson et al. (2007), we comprehensively summarize previously published meta-analyses on (a) the optimal and unit-weighted multiple correlations between the Big Five personality dimensions and behaviors in organizations, including job performance; (b) generalizable bivariate relationships of Conscientiousness and its facets (e.g., achievement orientation, dependability, cautiousness) with job performance constructs; (c) the validity of compound personality measures; and (d) the incremental validity of personality measures over cognitive ability. Hundreds of primary studies and dozens of meta-analyses conducted and published since the mid 1980s indicate strong support for using personality measures in staffing decisions. Moreover, there is little evidence that response distortion among job applicants ruins the psychometric properties, including criterion-related validity, of personality measures. We also provide a brief evaluation of the merits of alternatives that have been offered in place of traditional self-report personality measures for organizational decision making. Given the cumulative data, writing off the whole domain of individual differences in personality or all self-report measures of personality from personnel selection and organizational decision making is counterproductive for the science and practice of I-O psychology.

Personnel Psychology recently published the revised and enhanced transcripts from a panel discussion held during the 2004 annual conference

None of the authors is affiliated with any commercial tests, including personality, integrity, and ability tests.

Correspondence and requests for reprints should be addressed to Deniz S. Ones, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455-0344; deniz.s.ones-1@tc.umn.edu.

of the Society for Industrial and Organizational Psychology. The panelists included M. A. Campion, R. L. Dipboye, J. R. Hollenbeck, K. Murphy, and N. Schmitt who, among them, served as editors or associate editors of *Personnel Psychology* and *Journal of Applied Psychology* between 1989 and 2002. The panelists, and the article generated (hereafter referred to as Morgeson et al. [2007]) questioned the usefulness of personality measures in personnel selection primarily on the grounds of purportedly low validity and secondarily because of perceived problems with response distortion.

In this article, we first scrutinize and summarize the meta-analytic research documenting the usefulness of personality variables for a broad spectrum of organizational attitudes, behaviors, and outcomes. Second, we consider the impact of potential response distortion on the usefulness of personality measures. Third, we briefly evaluate the desirability, feasibility, and usefulness of alternatives that have been offered in place of self-report personality measures for organizational decision making.

What Are the Validities of Personality Measures?

Central to Morgeson et al.'s critique was the argument that the validity of personality measures is low and has changed little since Guion and Gottier's (1965) negative review. However, qualitative reviews often reach erroneous conclusions (Cooper & Rosenthal, 1980), and vote counting approaches to literature review, such as that employed by Campion in the target article (as well, of course, by Guion and Gottier's review), are similarly defective (e.g., Hedges & Olkin, 1980; Hunter & Schmidt, 2004).¹ In responding to claims by Morgeson et al., we rely on meta-analytic, quantitative summaries of the literature.²

¹Later in this article, using an example from Campion's table, we illustrate how such summaries lead to erroneous conclusions about the state of the research literature on personality and social desirability.

²Campion asserted that "gathering all of these low-quality unpublished articles and conducting a meta-analysis does not erase their limitations. . . . [T]he findings of the meta-analyses cannot be believed uncritically. I think they overestimate the criterion-related validity due to methodological weaknesses in the source studies" (p 707).

Appropriately conducted meta-analyses are not necessarily adversely affected by low-quality studies; in fact, meta-analysis is a quantitative way of testing the effect such low-quality studies have on conclusions drawn from a body of literature. Moderator analyses related to the quality of the sources, such as those now frequently conducted in meta-analyses, will reveal systematic shortcomings of low-quality studies (whether they inflate or deflate the magnitude, or simply increase the variability in observed effect sizes). Vote counting approaches, however, lack such systematic and objective mechanisms to safeguard us against low-quality studies unduly influencing conclusions drawn from literature reviews.

The Evidence: Multiple Correlations Based on Quantitative Reviews

We sought and located the most comprehensive meta-analyses that have examined the relationships between the Big Five and the following variables: (a) performance criteria (e.g., overall job performance, objective and task performance, contextual performance, and avoidance of counterproductive behaviors), (b) leadership criteria (emergence, effectiveness, and transformational leadership), (c) other criteria such as team performance and entrepreneurship, and (d) work motivation and attitudes.

In our summary we report both uncorrected, observed relationships, as well as corrected, operational validity estimates. Observed correlations are biased due to the influence of statistical artifacts, especially unreliability in the criteria and range restriction (see *Standards for Educational and Psychological Testing*; AERA, APA, & NCME, 1999). Thus, it is our position that conclusions should be based on operational (corrected) validities. This has been the scientifically sound practice in over a dozen meta-analyses of personality measures and in over 60 meta-analyses of cognitive and noncognitive personnel selection tools (e.g., interviews, assessment centers, biodata, situational judgment tests).

Morgeson et al. (2007) criticized some previous meta-analyses of personality variables on the grounds that corrections were applied for either predictor unreliability or imperfect (incomplete) construct measurement (or both), a procedure that “overestimate[s] the operational validity” (p. 706). For the summary that follows, we employed only values corrected for attenuation due to criterion unreliability (and range restriction where appropriate).³ We do wish to note that if one carefully scrutinizes the personnel selection literature, the meta-analyses on personality generally have made *more conservative* corrections—using higher estimates for the reliability of performance (Barrick & Mount, 1991) and not

³Operational validity estimates should only be corrected for attenuation due to unreliability in the criterion (and, where appropriate, range restriction), but not for unreliability in the predictor measures, in order to provide realistic estimates of the utility of tests as they are used in applied settings, including in selection (Hunter & Schmidt, 1990, 2004; Morgeson et al., 2007).

In cases where estimates presented in meta-analyses were corrected for predictor unreliability or imperfect construct measurement, we attenuated these estimates using the original artifact distributions reported in the meta-analyses. If no predictor reliabilities were reported, we employed meta-analytically derived estimates of personality scale reliabilities provided by Viswesvaran and Ones (2000). In meta-analyses where relationships were only corrected for sampling error (i.e., sample-size weighted mean *r*s), we applied corrections for attenuation due to unreliability in the criterion, using reliability estimates reported in the original meta-analyses where possible. In the few cases where criterion reliabilities were not available to make these corrections, we used appropriate reliability estimates from the literature (see notes to Tables 1, 2, and 3 for details). This strategy enabled us to summarize only psychometrically comparable validities that represent appropriate estimates of the applied usefulness of personality measures.

correcting for range restriction—unlike meta-analyses of many other kinds of predictors.

Some previous meta-analyses of personality have only reported bivariate relationships with criteria (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000). Morgeson et al. (2007) focused on such observed bivariate values. Other studies have used meta-analytic estimates of bivariate relationships to compute multiple correlations documenting the usefulness of the Big Five personality dimensions as a set (e.g., Judge, Bono, Ilies, & Gerhardt, 2002). In direct response to Morgeson et al.'s general assertion that personality is not useful, we focus on multiple correlations (both optimally weighted and unit-weighted) for the Big Five as a set. For meta-analyses that did not report a multiple correlation of the Big Five predicting the criterion of interest, operational validities reported for individual Big Five traits were used to compute multiple R s.⁴ At the typical primary study level, multiple regression weights are often problematic. However, basing multiple regression on meta-analytic intercorrelation matrices to a large extent circumvents capitalization on chance. Nevertheless, we computed and report unit-weighted composite correlations for each criterion as well.

Table 1 summarizes the operational usefulness of the Big Five as a set in predicting job performance and its facets, leadership, and other variables in organizational settings. The R s provided represent the best, meta-analytically derived estimates of the applied usefulness of Big Five personality measures in the prediction of work-related criteria.

For job performance criteria, optimally weighted R s, computed using meta-analytic intercorrelations among variables, range between .1 and .45. Individual overall job performance is predicted with $R = .27$. Objective (nonrating) measures of performance are predicted with $R = .23$. Note that personality variables measured at the Big Five level are even more predictive of performance facets such as counterproductive work behaviors (R s .44 and .45 for avoiding interpersonal and organizational deviance, respectively), organizational citizenship behaviors ($R = .31$), and interpersonal behaviors of getting along ($R = .33$) and individual teamwork ($R = .37$).

In addition to performance and facets of performance, the Big Five personality variables are far from trivial predictors of leadership and other criteria such as training performance. R s from optimally weighted composites range between .30 and .49 for leadership criteria. There are sizeable

⁴To estimate the operational validity of the Big Five as a set, operational validities were obtained from the most comprehensive meta-analyses available and used to compute optimally and unit-weighted multiple correlations between measures of the Big Five and each of the criteria. Meta-analytic intercorrelations among the Big Five (corrected only for sampling error) were obtained from Ones (1993; also available in Ones, Viswesvaran, & Reiss 1996).

TABLE 1
Meta-Analytic Validities of the Big Five for Predicting Work-Related Criteria

Criterion	Meta-analytic source	Observed		Operational	
		R unit	R optimal	R unit	R optimal
Job performance					
Overall job performance	Barrick, Mount, & Judge (2001)	.12 ^a	.14 ^a	.23 ^a	.27 ^a
Objective performance	Barrick, Mount, & Judge (2001)	.11 ^b	.13 ^b	.20 ^b	.23 ^b
Getting ahead	Hogan & Holland (2003)	.16 ^c	.18 ^c	.24 ^{cd}	.27 ^{cd}
Getting along	Hogan & Holland (2003)	.17 ^c	.20 ^c	.26 ^{cd}	.33 ^{cd}
Expatriate performance	Mol, Born, Willemsen, Van Der Molen (2005)	.25 ^e	.26 ^e	.35 ^e	.37 ^e
Individual teamwork	Barrick, Mount, & Judge (2001)	.23 ^b	.24 ^b	.35 ^b	.37 ^b
Contextual: Citizenship	Borman, Penner, Allen, & Motowidlo (2001)	.21 ^{efg}	.23 ^{efg}	.29 ^{efg}	.31 ^{efg}
Contextual: Altruism	Organ & Ryan (1995)	.11 ^{ef}	.12 ^{ef}	.11 ^{efh}	.13 ^{efh}
Contextual: Generalized compliance	Organ & Ryan (1995)	.17 ^{ef}	.19 ^{ef}	.20 ^{efh}	.23 ^{efh}
CWB: Interpersonal deviance	Berry, Ones, & Sackett (2007)	.29 ⁱ	.41	.32 ^{ij}	.44 ^j
CWB: Organizational deviance	Berry, Ones, & Sackett (2007)	.32	.40	.36 ^j	.45 ^j
Leadership criteria					
Leadership	Judge, Bono, Ilies, & Gerhardt (2002)	.30	.34	.39 ^k	.45 ^k
Leader emergence	Judge, Bono, Ilies, & Gerhardt (2002)	.31 ^l	.38 ^l	.40 ^k	.49 ^k
Leader effectiveness	Judge, Bono, Ilies, & Gerhardt (2002)	.28 ^l	.28 ^l	.36 ^k	.37 ^k
Transformational leadership	Bono & Judge (2004)	.24	.25	.28 ^j	.30 ^j
Other criteria					
Training performance	Barrick, Mount, & Judge (2001)	.19 ^b	.23 ^b	.33 ^b	.40 ^b
Entrepreneurship	Zhao & Seibert (2006)	.20 ^m	.31 ^m	–	–
Team performance	Peeters, Van Tuijl, Rutte, & Reymen (2006)	.36 ⁿ	.44 ⁿ	.49 ^{dkn}	.60 ^{dkn}
Motivational criteria					
Goal setting	Judge & Ilies (2002)	.18 ⁱ	.48	.21 ^{ik}	.54 ^k
Expectancy	Judge & Ilies (2002)	.17	.26	.23 ^k	.33 ^k
Self-efficacy	Judge & Ilies (2002)	.35	.38	.40 ^k	.45 ^k
Procrastination	Steel (2007)	.42 ⁱ	.64	.45 ^{dhi}	.70 ^{dh}

(Continued)

TABLE 1 (Continued)

Criterion	Meta-analytic source	Observed		Operational	
		<i>R</i> unit	<i>R</i> optimal	<i>R</i> unit	<i>R</i> optimal
Work attitudes					
Job satisfaction	Judge, Heller, & Mount (2002)	.28	.33	.31 ^k	.36 ^k
Career satisfaction	Ng, Eby, Sorensen, & Feldman (2005)	.31 ^o	.36 ^o	.33 ^h	.39 ^h

Note. All values are based on meta-analytic data. Observed multiple correlations (observed *R*) are based on observed intercorrelations among the Big Five personality dimensions and uncorrected mean predictor–criterion correlations. Operational multiple correlations (operational *R*) are based on observed intercorrelations among the Big Five personality dimensions and operational validity estimates (personality–criterion correlations corrected for unreliability in the criterion only, and range restriction where indicated). Meta-analytic intercorrelations for the Big Five were obtained from Ones (1993) and subsequently attenuated using reliability estimates from Viswesvaran and Ones (2000) in order to reflect sample size weighted mean correlations that do not take unreliability in the different predictor measures into account. Personality–criterion correlations were obtained from the meta-analytic sources listed above. If operational validity estimates were not available, mean observed values or population values were corrected appropriately using the reliability estimates as listed below. Unit-weighted composites were computed using composite theory formulas (see Hunter & Schmidt, 2004; Nunnally, 1978); optimally weighted composites were computed using multiple regression.

^aSecond-order meta-analytic validity estimates based on independent samples.

^bSecond-order meta-analytic validity estimates.

^cHogan Personality Inventory scale intercorrelations were obtained from the test manual.

^dOperational validity estimates in meta-analysis were also corrected for range restriction.

^eValidity estimates based on other-ratings of criteria only.

^fValidity for Emotional Stability, Extraversion, Agreeableness, and Conscientiousness only. Meta-analytic estimate for the validity of Openness was not available.

^gValidity estimates based on studies not included in Organ and Ryan (1995).

^hPersonality–criterion correlations on the population level, attenuated using predictor reliability estimates from Viswesvaran and Ones (2000).

ⁱInforming computation of the unit-weighted composite, negative predictor–criterion relationships were taken into account (disagreeableness for goal setting, introversion for interpersonal deviance, and lack of openness for procrastination).

^jMean sample-size weighted personality–criterion correlations, disattenuated using means of criterion reliability artifact distributions used in original meta-analysis.

^kPersonality–criterion correlations on the population level, attenuated using means of predictor reliability artifact distributions reported in original meta-analysis.

^lPersonality–criterion correlations on the population level, attenuated using means of predictor and criterion reliability artifact distributions reported in original meta-analysis.

^mMean sample-size weighted *d*-value on personality scales between regular managers and entrepreneurs, subsequently converted to a correlation coefficient (positive values indicating entrepreneurs scored higher on the personality scales).

ⁿEstimate for professional teams only. Based on correlations between team elevation on personality scales and team performance.

^oPersonality–criterion correlations on the population level, attenuated using predictor reliability estimates from Viswesvaran and Ones (2000) and criterion reliability estimate from Judge, Cable, Boudreau, and Bretz (1995).

Rs for training performance (.40) and entrepreneurship (.31) as well. Furthermore, work teams' elevations of the Big Five predict team level performance ($R = .60$).

Unit-weighted composite correlations tell a consistent and very similar story. The Big Five personality variables are predictive of job performance and its facets, leadership, and other work-related criteria. The unit-weighted *Rs* range between .11 and .49.

The summary provided in Table 1 certainly belies the view that the Big Five personality variables have little value for understanding and predicting important behaviors at work. Note that the meta-analyses summarized have relied on scores from traditional self-report personality measures of the sort criticized by the authors of the target article. Moreover, these values are not produced by magical or fanciful corrections. Indeed, they most certainly involve lower corrections than meta-analyses of other selection measures such as interviews, for example. Note that the observed *Rs* listed in Table 1 are also in the useful ranges. Observed *Rs* are on average only .08 correlational points smaller than operational *Rs* ($SD = .04$). In short, the table shows that self-report personality scale scores assessing the Big Five are useful for a broad spectrum of criteria and variables in organizational settings.

In Table 1, we also list relationships between the Big Five and other important work-related variables such as motivation and attitudes. These variables are listed for completeness sake and to provide a comprehensive overview of relevant variables that personality traits relate to. Although motivation and work attitudes are qualitatively different from performance criteria, they are related to important behaviors and outcomes such as the effort dimension of job performance and turnover (Campbell, Gasser, & Oswald, 1996). We realize that summarizing relationships between personality traits and these variables is not in keeping with some traditional views in industrial and organizational psychology, which consider solely performance measures as appropriate validation criteria. However, we believe that providing this summary is important in light of contemporary research that has expanded the traditional criterion domain. Readers who are merely interested in the validity of personality measures for predicting task or overall job performance may confine their evaluation to the respective analyses discussed above. Those readers who are interested in a more comprehensive list of outcomes related to personality may find the expanded list informative. As scientists, we cannot dictate what the goals of organizations should be and the types of workforces they want to create through their personnel selection practices. However, we can provide organizations with valuable information on the effectiveness of the selection tools that they have at their disposal, as well as the likely outcomes of their use in operational settings.

Validity of Conscientiousness and Its Facets

Because evidence suggests that Conscientiousness is the single best, generalizable Big Five predictor of job performance (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Hough & Ones, 2001; Schmidt & Hunter, 1998), Table 2 presents our summary of meta-analytic operational validities (bivariate relationships) for Conscientiousness and its facets in predicting dimensions of job performance. We summarized operational validity estimates for each predictor–criterion relationship from the most comprehensive meta-analyses to date.⁵ Operational validities for Conscientiousness range from $\rho = .04$ for the altruism facet of organizational citizenship behaviors to .38 for avoiding organizational deviance. Conscientiousness scales predict overall job performance ($\rho = .23$), objective performance indices ($\rho = .19$), and task performance ($\rho = .15$). Operational validities for interpersonal behaviors and aspects of contextual performance are in the .20s (e.g., getting along, organizational citizenship behaviors, individual teamwork).

The achievement facet of Conscientiousness predicts overall job performance ($\rho = .18$), task performance ($\rho = .22$), and job dedication aspect of contextual performance ($\rho = .34$) well. The dependability facet of Conscientiousness appears to be even more valuable for many criteria. Operational validities for dependability are .22 for overall job performance, .15 for task performance, .18 for interpersonal facilitation, .40 for job dedication, and .30 for avoiding counterproductive work behaviors.

The validities we have summarized for the Big Five (Table 1) and for Conscientiousness and its facets (Table 2) are on par with other valuable predictors frequently used in personnel selection and assessment. For example, meta-analytic operational validities for job performance have been reported as .36 for assessment centers (Gaugler, Rosenthal, Thornton, & Bentson, 1987), .44 for structured interviews (McDaniel, Whetzel, Schmidt, & Maurer, 1994), .35 for biodata (Rothstein, Schmidt, Erwin, Owens, & Sparks, 1990, as cited in Schmidt & Hunter, 1998), and .26 and .34 for situational judgment tests (McDaniel, Hartman, Whetzel, & Grubb, 2007; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). Most researchers and practitioners in I-O psychology regard the magnitude of these validities, which are comparable to those of personality variables (see Tables 1 and 2), as “useful” in applied settings (see McDaniel et al., 2001; Schmitt, Rogers, Chan, Sheppard, & Jennings,

⁵Again, validities listed are the values that have *not* been corrected for unreliability in the predictor or for imperfect construct measurement. If operational validity estimates were not reported, the appropriate corrections were applied to attenuate or disattenuate the reported values (details on each correction and the reliability estimates used are provided in the note to Table 2).

TABLE 2
*Operational Validity of Conscientiousness and Its Facets for
 Performance-Related Criteria*

Criterion/predictor	Meta-analytic source	\bar{r}	ρ
Job Performance			
Overall job performance			
Conscientiousness	Barrick, Mount, and Judge (2001)	.12 ^a	.23 ^a
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.10	.18 ^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.13	.22 ^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.05	.09 ^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.01	-.01 ^b
Objective performance			
Conscientiousness	Barrick, Mount, and Judge (2001)	.10 ^c	.19 ^c
Getting ahead			
Conscientiousness	Hogan and Holland (2003)	.12	.17 ^d
Getting along			
Conscientiousness	Hogan and Holland (2003)	.14	.21 ^d
Task performance			
Conscientiousness	Hurtz and Donovan (2000)	.10	.15
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.13	.22 ^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.09	.15 ^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.08	.14 ^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	.06	.10 ^b
Expatriate performance			
Conscientiousness	Mol, Born, Willemsen, Van Der Molen (2005)	.17 ^e	.24 ^e
Individual team work			
Conscientiousness	Barrick, Mount, and Judge (2001)	.15 ^e	.23 ^c
Contextual: OCB			
Conscientiousness	LePine, Erez, and Johnson (2002)	.19	.21 ^f
Contextual: Citizenship			
Conscientiousness	Borman, Penner, Allen, and Motowidlo (2001)	.19 ^g	.26 ^g
Contextual: Altruism			
Conscientiousness	Organ and Ryan (1995)	.04 ^e	.04 ^e
Contextual: Generalized compliance			
Conscientiousness	Organ and Ryan (1995)	.17 ^e	.21 ^e
Contextual: Interpersonal facilitation			
Conscientiousness	Hurtz and Donovan (2000)	.11	.16
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.06	.10 ^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.11	.18 ^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.01	-.02 ^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	.00	.00 ^b
Contextual: Job dedication			
Conscientiousness	Hurtz and Donovan (2000)	.12	.18
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.20	.34 ^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.23	.40 ^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.05	.09 ^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	.04	.07 ^b

(Continued)

TABLE 2 (Continued)

Criterion /predictor	Meta-analytic source	\bar{r}	ρ
CWB: Overall counterproductivity			
Conscientiousness	Berry, Ones, and Sackett (2007)	-.28 ^h	-.31 ^b
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.00	.00 ^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.21	-.30 ^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.04	-.06 ^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.06	-.10 ^b
CWB: Interpersonal deviance			
Conscientiousness	Berry, Ones, and Sackett (2007)	-.19	-.21 ^b
CWB: Organizational Deviance			
Conscientiousness	Berry, Ones, and Sackett (2007)	-.34	-.38 ^b
Leadership criteria			
Leadership			
Conscientiousness	Judge, Bono, Ilies, and Gerhardt (2002)	.20	.26 ^b
Leader emergence			
Conscientiousness	Judge, Bono, Ilies, and Gerhardt (2002)	.23 ^h	.30 ^b
Leader effectiveness			
Conscientiousness	Judge, Bono, Ilies, and Gerhardt (2002)	.11 ^h	.15 ^b
Transformational Leadership			
Conscientiousness	Judge, Bono, Ilies, and Gerhardt (2002)	.10	.12 ^b
Other criteria			
Training performance			
Conscientiousness	Barrick, Mount, and Judge (2001)	.13 ^c	.23 ^c
Entrepreneurship			
Conscientiousness	Zhao and Seibert (2006)	.19 ⁱ	—
Team's Performance			
Conscientiousness	Peeters, Van Tuijl, Rutte, and Reymen (2006)	.29 ^j	.40 ^{bdj}
Motivation criteria			
Goal setting			
Conscientiousness	Judge and Ilies (2002)	.22	.26 ^b
Expectancy			
Conscientiousness	Judge and Ilies (2002)	.16	.21 ^b
Self-efficacy			
Conscientiousness	Judge and Ilies (2002)	.17	.20 ^b
Procrastination			
Conscientiousness	Steel (2007)	-.62	-.68 ^{df}
Work attitudes			
Job satisfaction			
Conscientiousness	Judge, Heller, and Mount (2002)	.20	.22 ^b
Career satisfaction			
Conscientiousness	Ng, Eby, Sorensen, and Feldman (2005)	.12 ^k	.13 ^f

(Continued)

TABLE 2 (Continued)

Criterion /predictor	Meta-analytic source	\bar{r}	ρ
Learning criteria			
Motivation to learn			
Conscientiousness	Colquitt, LePine, and Noe (2000)	.31	.35 ^f
Achievement motivation	Colquitt, LePine, and Noe (2000)	.27	.31 ^l
Skill acquisition			
Conscientiousness	Colquitt, LePine, and Noe (2000)	-.04	-.05 ^f
Achievement motivation	Colquitt, LePine, and Noe (2000)	.13	.15 ^l

Note. All values are based on meta-analytic data. Personality–criterion correlations were obtained from the meta-analytic sources listed above. If operational validity estimates were not available, mean observed values or population values were corrected appropriately using the reliability estimates listed below.

\bar{r} = observed validity (uncorrected, sample-size weighted mean correlation). ρ = operational validity (corrected for unreliability in the criterion only, and range restriction where indicated).

^aSecond-order meta-analytic validity estimates based on independent samples.

^bPersonality–criterion correlation on the population level, attenuated using mean of predictor reliability artifact distribution reported in original meta-analysis.

^cSecond-order meta-analytic validity estimates.

^dOperational validity estimates in meta-analysis were also corrected for range restriction.

^eOperational validity estimate based on other-ratings of criteria only.

^fPersonality–criterion correlations on the population level, attenuated using predictor reliability estimates from Viswesvaran and Ones (2000).

^gOperational validity estimate based on studies not included in Organ and Ryan (1995).

^hPersonality–criterion correlations on the population level, attenuated using means of predictor and criterion reliability artifact distributions reported in original meta-analysis.

ⁱMean sample-size weighted *d*-value on personality scales between regular managers and entrepreneurs, subsequently converted to a correlation coefficient (positive values indicating entrepreneurs scored higher on the personality scales).

^jEstimate for professional teams only. Based on correlations between team elevation on personality scales and team performance.

^kPersonality–criterion correlation on the population level, attenuated using predictor reliability estimates from Viswesvaran and Ones (2000) and criterion reliability estimate from Judge, Cable, Boudreau, and Bretz (1995).

^lPersonality–criterion correlations on the population level, attenuated using predictor reliability estimates from Connolly and Ones (2007a).

1997). If the criticism that the amount of variance explained by personality is supposedly trivial was to be considered seriously, then costly and time-consuming assessment tools such as situational judgment tests, assessment centers, structured interviews, and biodata do not fare better than traditional self-report measures of personality.

Different Sets of Personality Variables Are Useful for Different Occupational Groups

There is a strong tradition of using ability-based predictors in personnel selection (see Ones, Viswesvaran, & Dilchert, 2004, for those meta-analyses not listed in Campion's Table 2). A key finding in the

cognitive ability literature is that regardless of the occupation or job under consideration, ability tests predict overall job performance (Ones, Viswesvaran, & Dilchert, 2005a), and there appears to be little predictive gain from specific abilities beyond general cognitive ability (Brown, Le, & Schmidt, 2006; Ree, Earles, & Teachout, 1994). This is one aspect in which the domains of ability and personality differ. Apart from Conscientiousness, there seem to be no other personality traits that predict overall job performance with similarly consistent validities across different jobs. Instead, for different occupations, different combinations of the Big Five yield the best levels of validity.

Table 3 presents a summary of meta-analytic personality validity estimates in predicting job performance among various occupational groups.⁶ Operational validities that exceed .10 have been highlighted in the Table. For professionals, only Conscientiousness scales appear to be predictive of overall job performance. Similarly, for sales jobs, only Conscientiousness and its facets of achievement, dependability, and order predict overall performance well. For skilled and semi-skilled jobs, in addition to Conscientiousness, Emotional Stability appears to predict performance. For police and law enforcement jobs, Conscientiousness, Emotional Stability, and Agreeableness are useful personal characteristics. In customer service jobs, all the Big Five dimensions predict overall job performance. Finally, for managers, the Extraversion facets dominance and energy, and the Conscientiousness facets achievement and dependability, are predictive. Thus, different sets of personality variables are useful in predicting job performance for different occupational groups. The magnitudes of operational validities of single valuable Big Five dimensions for specific occupations range between .11 and .25.

Compound Personality and Other Traits

We should note that the Big Five do not exhaust the set of useful personality measures. Judge, Locke, and colleagues' concept of core self-evaluations has shown itself a useful predictor of performance, controlling for Conscientiousness (Erez & Judge, 2001) or the entire set of Big Five traits (Judge, Erez, Bono, & Thoresen, 2003). Crant and colleagues' measure of proactivity has similarly shown itself to be useful, even when controlling for the entire set of Big Five traits (Crant, 1995). It may be that these measures pick up more complex personality traits.

Personality scales that assess more than one dimension of the Big Five are referred to as compound personality scales (cf. Hough & Ones, 2001;

⁶Footnote 5 applies here as well.

TABLE 3
Personality Traits Predictive of Job Performance for Various Occupational Groups

Occupation	Meta-analytic source	\bar{r}	ρ
Sales			
Emotional Stability	Barrick, Mount, and Judge (2001)	.03 ^a	.05 ^a
Extraversion	Barrick, Mount, and Judge (2001)	.07 ^a	.09 ^a
Openness	Barrick, Mount, and Judge (2001)	-.01 ^a	-.02 ^a
Agreeableness	Barrick, Mount, and Judge (2001)	.05 ^a	.01 ^a
Conscientiousness	Barrick, Mount, and Judge (2001)	.11 ^a	.21^a
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.15	.26^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.13	.23^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.10	.17^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.02	-.04 ^b
Skilled or semi-skilled			
Emotional Stability	Barrick and Mount (1991); Hurtz and Donovan (2000); Salgado (1997)	.06 ^c	.11^c
Extraversion	Barrick, Mount, and Judge (2001)	.03 ^a	.05 ^a
Openness	Barrick, Mount, and Judge (2001)	.03 ^a	.04 ^a
Agreeableness	Barrick, Mount, and Judge (2001)	.05 ^a	.08 ^a
Conscientiousness	Barrick, Mount, and Judge (2001)	.12 ^a	.19^a
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.11	.18^b
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.14	.24^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.11	.19^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.11	-.18^b
Customer service			
Emotional Stability	Hurtz and Donovan (2000)	.08	.12
Extraversion	Hurtz and Donovan (2000)	.07	.11
Openness	Hurtz and Donovan (2000)	.10	.15
Agreeableness	Hurtz and Donovan (2000)	.11	.17
Conscientiousness	Hurtz and Donovan (2000)	.17	.25
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.12	.20^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	.07	.11^b
Professional			
Emotional Stability	Barrick, Mount, and Judge (2001)	.04 ^a	.06 ^a
Extraversion	Barrick, Mount, and Judge (2001)	-.05 ^a	-.05 ^a
Openness	Barrick, Mount, and Judge (2001)	-.05 ^a	-.08 ^a
Agreeableness	Barrick, Mount, and Judge (2001)	.03 ^a	.05 ^a
Conscientiousness	Barrick, Mount, and Judge (2001)	.11 ^a	.20^a
Police			
Emotional Stability	Barrick, Mount, and Judge (2001)	.07 ^a	.11^a
Extraversion	Barrick, Mount, and Judge (2001)	.06 ^a	.06 ^a
Openness	Barrick, Mount, and Judge (2001)	.02 ^a	.02 ^a
Agreeableness	Barrick, Mount, and Judge (2001)	.06 ^a	.10^a
Conscientiousness	Barrick, Mount, and Judge (2001)	.13 ^a	.22^a

(Continued)

TABLE 3 (Continued)

Occupation	Meta-analytic source	\bar{r}	ρ
Managerial			
Emotional Stability	Barrick, Mount, and Judge (2001)	.05 ^a	.08 ^a
Extraversion	Barrick, Mount, and Judge (2001)	.10 ^a	.17^a
Dominance	Hough, Ones, and Viswesvaran (1998)	.16	.27
Sociability	Hough, Ones, and Viswesvaran (1998)	-.01	-.02
Energy	Hough, Ones, and Viswesvaran (1998)	.12	.20
Openness	Barrick, Mount, and Judge (2001)	.05 ^a	.07 ^a
Flexibility	Hough, Ones, and Viswesvaran (1998)	.05	.11
Intellectance	Hough, Ones, and Viswesvaran (1998)	.07	.06
Agreeableness	Barrick, Mount, and Judge (2001)	.04 ^a	.08 ^a
Conscientiousness	Barrick, Mount, and Judge (2001)	.12 ^a	.21^a
Achievement	Hough, Ones, and Viswesvaran (1998)	.09	.17
Achievement	Dudley, Orvis, Lebiecki, and Cortina (2006)	.07	.12^b
Dependability	Hough, Ones, and Viswesvaran (1998)	.02	.03
Dependability	Dudley, Orvis, Lebiecki, and Cortina (2006)	.10	.17^b
Order	Dudley, Orvis, Lebiecki, and Cortina (2006)	-.06	-.11^b
Cautiousness	Dudley, Orvis, Lebiecki, and Cortina (2006)	.01	.01 ^b

Note. All values are based on meta-analytic data. Personality-criterion correlations were obtained from the meta-analytic sources listed above. If operational validity estimates were not available, mean observed values or population values were corrected appropriately using the reliability estimates listed below.

\bar{r} = observed validity (uncorrected, sample-size weighted mean correlation). ρ = operational validity (corrected for unreliability in the criterion only and range restriction where indicated).

^aSecond-order meta-analytic validity estimate.

^bPersonality-criterion correlation on the population level, attenuated using mean of predictor reliability artifact distribution reported in original meta-analysis.

^cSecond-order meta-analytic validity estimate. Because an estimate of this personality-criterion relationship was not available in Barrick et al. (2001), validity estimates from these sources were sample-size weighted and then averaged. For the operational validity estimate, the population level correlation reported in Barrick and Mount (1991) was attenuated using the Emotional Stability reliability estimate obtained from Viswesvaran and Ones (2000).

Hough & Oswald, 2000). Compound personality scales include integrity tests (see Ones & Viswesvaran, 2001a; Ones, Viswesvaran, & Schmidt, 1993), customer service scales (see Frei & McDaniel, 1998), drug and alcohol, stress tolerance, and violence scales (see Ones & Viswesvaran, 2001b), as well as managerial potential scales (e.g., Gough, 1984; see also Hough, Ones, & Viswesvaran, 1998; Ones, Hough, & Viswesvaran, 1998).

Perhaps the most studied of all compound personality traits is integrity. Regarding the predictive power and applied value of integrity tests, we wholeheartedly agree with N. Schmitt (see p. 695) that one should not "mislead clients with 'inflated' estimates of predictive validity using studies that are based on self-report criteria or concurrent studies." In fact, one should not mislead anyone (practitioners or researchers) with any

type of validity estimate. That is why we will take the opportunity to clarify the actual operational validity of integrity tests. In the meta-analysis conducted by Ones et al. (1993), the authors very carefully sorted out studies that were conducted using external criteria and predictive validation designs. They found enough studies to evaluate the validity of such measures against *non-self-report* broad counterproductive behaviors, overall job performance, and objective production records of performance. The operational validity of integrity tests for productivity (from production records) was .28 (see Ones et al.'s table 5). In *predictive* studies conducted on *job applicants*, the operational validity for *externally detected* counterproductive behaviors was .39 for overt integrity tests and .29 for personality-based integrity tests (see Ones et al.'s Table 11). Externally detected broad counterproductive behaviors for these two meta-analyses included violence on the job, absenteeism, tardiness, and disruptive behaviors but excluded theft (see footnote to Ones et al.'s Table 11). Finally, in *predictive* studies conducted on *job applicants*, the operational validity for *supervisory ratings of overall job performance* was .41 (see Ones et al.'s Table 8). These are the operational validities that organizations should keep in mind when integrity tests are considered for inclusion in personnel selection systems.

The criterion-related validities of other compound personality scales for predicting overall job performance have recently been summarized by Ones, Viswesvaran, and Dilchert (2005b). Ones and colleagues based their review on meta-analyses and focused on operational validities that were not corrected for predictor unreliability or imperfect construct measurement. The operational validities for compound scales other than integrity ranged between .19 and .42 for supervisory ratings of overall job performance. In fact, Ones et al., in their summary of the available meta-analyses, concluded that "not all personality traits are created equal in terms of their predictive and explanatory value. Highest validities for predicting overall job performance using predictors from the personality domain are found for compound personality variables. . . . The criterion-related validities of the compound personality variables presented . . . are among the highest for individual differences traits predicting overall job performance. In fact, only general mental ability has superior criterion-related validities for the same criterion." (p. 396).

Incremental Validity of Personality Variables

A contribution acknowledged by Murphy and Campion is the potential for personality inventories to provide incremental validity over cognitive variables. Nonetheless, Campion has suggested that previous reviews (specifically, Schmidt & Hunter, 1998) "overestimated" (p. 707)

the incremental validity for personality scales, on the grounds that criterion-related validities informing incremental validity calculations for Conscientiousness were inappropriately corrected for imperfect predictor construct measurement. Even if we put aside incremental validity computations for Conscientiousness, there is still evidence for substantial incremental validity of compound personality measures.

For integrity tests, the multiple correlations for predicting overall job performance using a measure of cognitive ability and integrity are reported in Table 12 of Ones et al. (1993). The values were .64 for the lowest complexity jobs, .67 for medium-complexity jobs, and .71 for the highest complexity jobs. Schmidt and Hunter (1998) updated the Ones et al. (1993) integrity test incremental validities for job performance by using a more precise estimate of the relationship between integrity test and cognitive ability test scores from Ones' (1993) meta-analysis. In the prediction of *overall job performance*, integrity tests yield .14 incremental validity points over tests of general mental ability.

Incremental validities of other compound personality scales have been reported in Ones and Viswesvaran (2001a; 2001b; in press). Incremental validities over cognitive ability in predicting overall job performance range between .07 and .16. It is important to note that none of these incremental validity computations involved corrections for incomplete construct measurement or unreliability in the predictor. There is incremental validity to be gained from using personality measures in predicting overall job performance, even when cognitive variables are already included in a given selection system.

Conclusions on the Usefulness of Personality Measures in Organizational Decision Making

The criterion-related validities of self-report personality measures are substantial. The Big Five personality variables as a set predict important organizational behaviors (e.g., job performance, leadership, and even work attitudes and motivation). The effect sizes for most of these criteria are moderate to strong (.20 – .50). Conscientiousness and its facets are predictive of performance constructs with useful levels of validity. Compound personality measures (e.g., integrity tests, customer service scales) are strong predictors of job performance (e.g., most operational validities for supervisory ratings of overall job performance are around .40), and they provide substantial incremental validity over tests of cognitive ability. The accumulated evidence supports the use of self-report personality scales in organizational decision making, including personnel selection.

Does Response Distortion Among Job Applicants Render Personality Measures Ineffective?

To determine whether response distortion influences psychometric properties of self-report personality measures, a critical evaluation of the faking, social desirability, and impression management literatures is necessary. There are two empirical approaches that have been used to examine effects on scale scores: lab studies of faking and studies aiming to examine response distortion among actual job applicants. It is crucial to distinguish between these two types of studies in reaching correct conclusions about the effect of response distortion on the psychometric properties of noncognitive measures. Lab studies of directed faking, where participants (mostly students) are either directed to fake good or respond as though they were job applicants, create direct experimental demands on participants. A key question is whether naturally occurring response distortion is adequately modeled in lab studies of directed faking (cf. Dilchert, Ones, Viswesvaran, & Deller, 2006).

Lab studies of directed faking typically find that criterion-related validity and construct validity of self-report personality scales are affected by instructions to fake. Directed faking studies can, by design, only document maximal limits on direct, outright lying by modeling response distortion upon strong experimental instructions. However, as Ruch and Ruch (1967, p. 201) noted, “the real-life situation is not the same as an all-out attempt to fake when subjects are instructed to do so as part of an experiment ‘that doesn’t count’ (Dunnette McCartney, Carlson, & Kirchner 1962).” There is emerging consensus among those who study faking and response distortion that faking demand effects in lab studies document a different phenomenon than response distortion among job applicants. Smith and Ellingson (2002, p. 211) pointed out that “most research demonstrating the adverse consequences of faking for construct validity uses a fake-good instruction set. The consequence of such a manipulation is to exacerbate the effects of response distortion beyond what would be expected under realistic circumstances (e.g., an applicant setting).”

Much of the misunderstandings in the literature stem from combining and/or confusing directed studies of faking and naturally occurring applicant response distortion (see, for example, Table 1 of Morgeson et al. [2007], where the problem is further compounded by a vote counting approach to literature review where large-scale meta-analyses are given the same weight as investigations conducted in extremely small samples). Accordingly, in the next sections of this article we examine several commonly voiced concerns regarding faking and social desirability. Given the interpretational problems with lab studies of directed faking, we confine our examination only to empirically supported conclusions based on actual job applicant data.

(Non) Effects of Faking and Social Desirability on Criterion-Related Validity

In Morgeson et al. (2007), Campion (Table 1) lists eight studies as providing support for the position that response distortion affects the validity of personality measures. An analysis of these eight studies is presented in Appendix A. Scrutinizing the eight studies listed indicates that five of the eight were studies of directed faking or lab studies where participants were assigned to experimental response distortion groups. Among these studies, Dunnette et al. (1962) investigated both “fakability” and the degree of actual faking in a selection context, and were very careful to conclude that although scores were increased when test-takers were instructed to fake, few applicants actually tend to fake in real employment situations. Another study offered as supporting the claim that criterion-related validity is affected by faking was conducted on participants in nonmotivating contexts. Only two of the eight studies were conducted on job applicants and only one of those two actually included a criterion for which criterion-related validity could be computed. In that study, criterion-related validities of personality scales were affected only after corrections for response distortion were applied. Thus, the only study that offered a criterion-related validity estimate computed among job applicants showed sizeable validity for the personality measure (but did not support the correction of scores for presumed faking).

The best method for estimating operational validities in selection settings is predictive validation based on applicants. In selection settings, job applicant groups should be the focal interest if response distortion is regarded as a concern. Two large-scale quantitative reviews have reported substantial criterion-related validities of personality measures in *predictive* studies, among *job applicants*. Hough (1998b) reported a summary of predictive validity studies for personality variables based on the Hough (1992) database (see Hough, 1998b, Table 9.6). Hough did not correct the mean observed validities she obtained for range restriction or unreliability in the criterion and predictor. Nevertheless, there were useful and sizeable predictive validities reported. Achievement measures predicted both job proficiency and training success with a mean observed validity of .19. Predictive studies for educational success yielded mean observed validities in the .15 to .23 range for potency (dominance facet of Extraversion), the achievement facet of Conscientiousness, and Emotional Stability (adjustment). Finally, predictive validity studies for the counterproductive work behaviors criterion found substantial mean observed validities for the achievement facet of Conscientiousness (–.33), the dependability facet of Conscientiousness (–.23), Emotional Stability (–.17), and intellectance (.24).

As mentioned earlier, Ones et al. (1993) reported the criterion-related validities for integrity tests based only on predictive validation studies conducted on job applicants. The criterion-related validity of integrity tests is substantial, even in the potential presence of naturally occurring response distortion among job applicants (.41 for predicting overall job performance; .39 for overt tests predicting externally detected counterproductive behaviors; and .29 for personality-based integrity tests for externally detected counterproductive behaviors).

These quantitative reviews summarize predictive studies, where job applicants are tested and selected on the basis of their scores on personality measures. They indicate that there is useful criterion-related validity for self-report personality measures when the stakes involve getting a job.

(Non) Effects of Faking and Social Desirability on Construct Validity

Primary construct validity studies have focused on convergent/divergent validities of personality scales and/or factor structure under conditions that differ in the degree of response distortion expected. Findings in this domain show that factor structures of personality inventories hold up in job applicant samples (e.g., see Marshall, De Fruyt, Rolland, & Bagby, 2005, for the NEO PI-R; Smith & Ellingson, 2002; Smith, Hanges, & Dickson, 2001, for the Hogan Personality Inventory). That is, when the focus is on actual job applicants, factor structures underlying responses on personality inventories are in line with those theoretically postulated for the traits assessed by such inventories.

A recent meta-analytic investigation of divergent validity and factor structure of two widely used Big Five inventories (NEO PI-R and Hogan Personality Inventory, *N*s ranging from 1,049 to 10,804) also indicated that the factor structures of the two inventories held up among job applicants (Bradley & Hauenstein, 2006). Contexts that have been postulated to motivate job applicants to distort their responses appear to have “a small, practically unimportant, influence on either the intercorrelations among the Big Five, and no effect on the higher order factor loadings of Big Five measures” (Bradley & Hauenstein, 2006, p. 328).

Investigations employing IRT analyses lead to equivocal conclusions. Although Stark, Chernyshenko, Chan, Lee, and Drasgow (2001) reported differential item functioning for applicants versus nonapplicants on a personality test, Robie, Zickar, and Schmit's (2001) analyses suggest that “personality measures used for selection retain similar psychometric properties to those used in incumbent validation studies.”

A recently published investigation using a within-subjects design (the same individuals completing a personality measure once under selection,

and once under development conditions) concluded that there was only a limited degree of response distortion (Ellingson, Sackett, & Connelly, 2007). As both factor structure and convergent/divergent validities hold up in job applicant samples, potential response distortion among job applicants does not appear to destroy the construct validity of personality measures. There is strong evidence that self-report personality measures “can be successfully used to capture traits in motivating contexts” (Smith & Ellingson, 2002, p. 217).

Presumed Usefulness of Social Desirability Scales in Predicting Various Job Relevant Criteria

Some authors have suggested (cf. Marcus, 2003, see also Murphy, p. 712) that impression management may be a useful personal characteristic and even possess predictive validity in occupational settings. Yet, meta-analyses and large sample primary studies of social desirability scales report negligibly small correlations for overall job performance in heterogeneous samples (Ones, Viswesvaran, & Reiss, 1996) among managers (Viswesvaran, Ones, & Hough, 2001) and among expatriates (Foldes, Ones, & Sinangil, 2006). In a meta-analysis that separated self-deceptive enhancement from impression management aspects of social desirability, Li and Bagger (2006) also found weak support for the criterion-related validity of these scales: Neither aspect of social desirability predicted performance well. Social desirability, self-deceptive enhancement, and impression management scales do not appear to be useful in predicting job performance or its facets.

Conclusion: The Impact of Response Distortion Among Job Applicants Is Overrated

Understandably, in personnel selection settings, faking and response distortion are a concern on any self-report, noncognitive measure (be it a personality test, an interview, a biodata form, or a situational judgment test with behavioral tendency instructions). For personality measures, there is evidence from job applicant samples documenting their criterion-related validity and construct validity in motivating contexts, a fact that should alleviate some of these concerns.

An Evaluation of Alternatives

There are a number of alternatives offered by the authors of the target article that I-O psychologists can pursue with regard to assessing personality

in organizational settings. In this section, we evaluate the desirability, feasibility, and usefulness of each.

The Use of Faking Indices to Correct Scores

Correcting personality scale scores based on faking or social desirability scale scores has a long history in psychological assessment. Traditionally, social desirability, lie, or impression management scales are used in the corrections (Goffin & Christiansen, 2003). Research to date has failed to support the efficacy of such corrections in dealing with intentional response distortion (Christiansen, Goffin, Johnston, & Rothstein, 1994; Ellingson, Sackett, & Hough, 1999; Hough, 1998a). The cumulative research evidence shows that “if selection researchers are interested only in maximizing predicted performance or validity, the use of faking measures to correct scores or remove applicants from further employment consideration will produce minimal effects” (Schmitt & Oswald, 2006, p. 613).

Conditional Reasoning Measures of Personality

Individuals can distort their scores on personality measures when directly instructed to do so. In an effort to construct less fakable measures, L. James and colleagues have developed conditional reasoning tests of some behavioral tendencies, most prominently aggression (James et al., 2005). The premise behind conditional reasoning tests is that individuals use various justification mechanisms to rationalize their behaviors. It is postulated that test takers’ standing on a given trait is related to the justification mechanisms they endorse. Conditional reasoning measures are described as tapping into implicit personality characteristics, those characteristics and tendencies that individuals are not explicitly aware of.

Because of their covert way of assessing noncognitive characteristics, conditional reasoning measures have been suggested as substitutes (James et al., 2005) and supplements (Bing et al., 2007) to traditional self-report personality measures. Hollenbeck (p. 718) states, “I think conditional reasoning might be another approach that is really promising in this regard. These tests are less direct than traditional measures and focus on how people solve problems that appear, on the surface at least, to be standard inductive reasoning items. . . . These measures have been found to be psychometrically sound in terms of internal consistency and test–retest estimates of reliability, and their predictive validity against behavioral criteria, uncorrected, is over .40.” It is true that James et al. (2005) have presented evidence that documents the usefulness of conditional reasoning measures. However, it appears that the criterion-related validity was

drastically overestimated (Berry, Sackett, & Tobares, 2007). In fact, the cumulative evidence to date indicates that validities of the conditional reasoning test for aggression range between .15 and .24 for counterproductive work behaviors, whereas the mean validity for job performance is .14 (meta-analytic mean r ; Berry et al., 2007). Conditional reasoning tests have at best comparable validities to self-report personality measures, along with sizeable comparative disadvantages, owing to their difficulty and cost in development. Furthermore, additional research is needed on the incremental validity of conditional reasoning tests over ability measures in unrestricted samples as well as on group differences, adverse impact, and differential prediction.

Using Forced Choice Personality Inventories With Ipsative Properties

An approach that has been suggested as having the potential to prevent faking and therefore enhance the usefulness of self-report personality scales is using forced choice response formats. Forced choice measures ask test takers to choose among a set of equally desirable options loading on different scales/traits (e.g., Rust, 1999). Unfortunately, the use of forced choice response formats often yields fully or at least partially ipsative scores. Such scores pose severe psychometric problems, including difficulties in reliability estimation and threats to construct validity (see Dunlap & Cornwell, 1994; Hicks, 1970; Johnson, Wood, & Blinkhorn, 1988; Meade, 2004; Tenopyr, 1988). Furthermore, ipsative data suffer from artificially imposed multicollinearity, which renders factor analyses and validity investigations using such data problematic. Individuals cannot be compared normatively on each scale, rendering ipsative scores useless for most purposes in organizational decision making. In short, forced choice measures are psychometrically inappropriate for between-subjects comparison, which of course is the whole point in selection contexts.

Even though recent studies have suggested that some forced choice formats can yield less score inflation than normative scales under explicit instructions to fake (Bowen, Martin, & Hunt, 2002; Christiansen, Burns, & Montgomery, 2005), there is also evidence that suggests that a multidimensional forced choice format "is not a viable method to control faking" (Heggestad, Morrison, Reeve, & McCloy, 2006, p. 9). In Heggestad et al.'s individual-level analyses, the forced choice measure was as affected by directions to distort responses as when a Likert-type response scale was employed. Some researchers have reported higher validities for forced choice measures (e.g., Villanova, Bernardin, Johnson, & Dahmus, 1994). However, the precise reasons for the higher correlations are unclear. Interestingly, unlike traditional Likert-type personality scale scores, forced choice personality scores appear to correlate with cognitive ability when

individuals respond as job applicants (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006).

Overall, forced choice measures, even if they yield only partially ipsative scores, are not certain to improve criterion-related validity of personality measures (or to do so without compromising the underlying constructs assessed). The challenge for future research will be to find ways of recovering normative scores from traditional forced choice inventories. Recently, there have been promising IRT-based developments in this area (McCloy, Heggstad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & Williams, 2006) suggesting that normative scores could potentially be estimated from ipsative personality scales (Chernyshenko et al., 2006; Heggstad et al., 2006). However, the usefulness of these approaches in operational selection settings has not been tested and their impact on predictive validity is yet unknown.

Constructing One's Own Personality Measure

To the question "What recommendations would you give about the use of personality tests in selection contexts?" Schmitt responded "first, avoid published personality measures in most instances. The only time I would use them is if they are directly linked in a face valid way to some outcome. Second, I would construct my own measures that are linked directly to job tasks in a face-valid or relevant fashion" (p. 715).

The primary concern highlighted by Morgeson et al. (2007) is the purportedly low validity of personality scales. It is unclear to the present set of authors why homegrown scales should necessarily have superior validity to professionally constructed, well-validated measures that have been shown to possess generalizable predictive power. There is no evidence that we can unearth that suggests that freshly constructed personality scales for each individual application perform better (i.e., have higher criterion-related validity) than existing, professionally developed and validated off-the-shelf personality scales. On the contrary, there is reason to believe that homegrown measures can be deficient, contaminated, or both, and can be lacking in terms of construct and criterion-related validity unless extraordinary effort and resources are devoted to scale construction, refinement, and validation. Although we are not opposed to the idea of constructing customized personality measures for particular applications (e.g., organizations, jobs), we are also realistic about what it takes to develop personality measures to match the psychometric properties of well developed, off-the-shelf inventories. Harrison Gough, the author of the California Psychological Inventory, has often reminded us (most recently, Gough, 2005) that many items, which, on the surface, suggest to measure a given personality construct, often do not survive rigorous

empirical tests of scale construction. For every item that has been included on professionally developed personality scales, dozens of personality items have ended up in item graveyards.

Again, if the issue is poor predictive validity of self-report measures, which Campion, Murphy, and Schmitt suggest, then constructing a new measure for each local application is unlikely to provide much improvement. Such homegrown scales may or may not have construct validity problems. Their item endorsement rates (i.e., the equivalent of item difficulty for the personality domain) may or may not be well calibrated. Unlike well developed existing personality measures, they may not have extensive norms for use in decision making. In short, there are no guarantees that homegrown personality measures would be psychometrically superior; in fact, it is reasonable to expect that the opposite may be the case.

Others' Ratings of Personality in Personnel Decision Making

An alternative to self-report measures of personality is to seek alternate rating sources. In organizational decision making, peers and previous supervisors could provide personality ratings, especially for internal candidates. Measuring constructs using multiple measures is always a good idea, not only in the domain of personality. There are off-the-shelf personality inventories with established psychometric characteristics that have been specifically adapted for use with others' ratings (e.g., NEO-PI-R).

A few field studies have examined the criterion-related validity of peer ratings of personality for job performance (Mount, Barrick, & Strauss, 1994), as well as high school grades (Bratko, Chamorro-Premuzic, & Saks, 2006) and career success (Judge, Higgins, Thoresen, & Barrick, 1999). Criterion-related validities appear to be somewhat higher than those of self-reports; other-ratings also have incremental validity over self-reports. Future research should continue to explore the potential for other-ratings of personality in personnel selection, placement, and promotion decisions.

Strategies to Enhance Prediction From Personality Variables

There is ongoing research into ways to enhance criterion-related validities of personality measures in applied settings. Promising ideas being pursued include nonlinear effects (e.g., for Emotional Stability, see Le, Ilies, & Holland, 2007), dark side traits (e.g., Benson & Campbell, 2007), profile analyses (Dilchert, 2007), and interactions between personality variables and other individual differences characteristics (see Judge & Erez, 2007, Witt, Burke, Barrick, & Mount, 2002). Another intriguing approach is to use multiple personality measures to assess the same trait. This

approach has been shown to yield up to 50% increase in criterion-related validity for Conscientiousness, depending on the number of scales combined (Connelly & Ones, 2007b), and has been shown to be independent of the effect of merely increasing the reliability of the measures. Future research should further pursue some of these approaches.

Abandoning the Use of Personality Measures in Personnel Selection and Decision Making

Morgeson et al. (2007) encourage I-O psychologists to reconsider the use of published self-report personality tests in personnel selection contexts. The authors state as the primary reason “the very low validity of personality tests for predicting job performance” (p. 683). We think that the quantitative summaries that we have reviewed in the first section of this article refute this assertion. Even validities of .20 translate to substantial utility gains. As reviewed above, taken as a set, the Big Five personality variables have multiple correlations in the .40s with many important organizational behaviors. Single compound traits also correlate around .40 with supervisory ratings of performance. Such measures can be expected to provide incremental validity and, thus, utility in applied settings.

If broad judgments are to be rendered about the validity of personality measures in personnel selection, we advise researchers to consider the multiple correlations of the measures in use, as well as validities of compound personality measures. To evaluate personality measures based on the Big Five as a set is hardly cheating, as complete inventories of the Big Five dimensions are, still, among the shortest selection measures available. Multiple correlations between the Big Five traits and many important organizational behaviors and outcomes are generally in the .30s and .40s, which surely ranks among the more important discoveries in personnel selection in the past quarter century. As we have shown in our summary, these relationships are substantial, as high as or higher than for other noncognitive measures.

In conclusion, “there is considerable evidence that both general cognitive ability and broad personality traits (e.g., Conscientiousness) are relevant to predicting success in a wide array of jobs” (Murphy & Shiarella, 1997, p. 825). The validities of noncognitive predictors, including personality inventories, are “practically useful” (Schmitt et al., 1997, p. 721). What is remarkable about self-report personality measures, beyond their criterion-related validity for overall job performance and its facets, is that they are also useful in understanding, explaining, and predicting significant work attitudes (e.g., job satisfaction) and organizational behaviors (e.g., leadership emergence and effectiveness, motivation and effort). Any theory of organizational behaviors that ignores personality variables would be

incomplete. Any selection decision that does not take the key personality characteristics of job applicants into account would be deficient.

Conclusions

1. Personality variables, as measured by self-reports, have substantial validities, which has been established in several quantitative reviews of hundreds of peer-reviewed research studies.
2. Vote counting and qualitative opinions are scientifically inferior alternatives to quantitative reviews and psychometric meta-analysis.
3. Self-reports of personality, in large applicant samples and actual selection setting (where faking is often purported to distort responses), have yielded substantial validities even for externally obtained (e.g., supervisory ratings, detected counterproductive behaviors) and/or objective criteria (e.g., production records).
4. Faking does not ruin the criterion-related or construct validity of personality scores in applied settings.
5. Other noncognitive predictors may derive their validity from capturing relevant personality trait variance.
6. Customized tests are not necessarily superior to traditional standardized personality tests.
7. When feasible, utilizing both self and observer ratings of personality likely produces validities that are comparable to the most valid selection measures.
8. Proposed palliatives (e.g., conditional reasoning, forced choice ipsative measures), when critically reviewed, do not currently offer viable alternatives to traditional self-report personality inventories.

REFERENCES

- Abbott RD, Harris L. (1973). Social desirability and psychometric characteristics of the Personal Orientation Inventory. *Educational and Psychological Measurement*, *33*, 427–432.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Barrick MR, Mount MK. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *PERSONNEL PSYCHOLOGY*, *44*, 1–26.
- Barrick MR, Mount MK, Judge TA. (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.
- Benson ML, Campbell JP. (2007, April). Emotional Stability: Inoculator of the dark-side personality traits/leadership performance link? In Ones DS (Chair). *Too much, too little, too unstable: Optimizing personality measure usefulness*. Symposium conducted

- at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York City, New York.
- Berry CM, Ones DS, Sackett PR. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*, 410–424.
- Berry CM, Sackett PR, Tobares V. (2007, April). *A meta-analysis of conditional reasoning tests of aggression*. Poster session presented at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York City, NY.
- Bing MN, Stewart SM, Davison H, Green PD, McIntyre MD, James LR. (2007). An integrative typology of personality assessment for aggression: Implications for predicting counterproductive workplace behavior. *Journal of Applied Psychology*, *92*, 722–744.
- Bono JE, Judge TA. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, *89*, 901–910.
- Borman WC, Penner LA, Allen TD, Motowidlo SJ. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, *9*, 52–69.
- Bowen C-C, Martin BA, Hunt ST. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis*, *10*, 240–259.
- Bradley KM, Hauenstein NMA. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science*, *48*, 313–335.
- Bratko D, Chamorro-Premuzic T, Saks Z. (2006). Personality and school performance: Incremental validity of self- and peer-ratings over intelligence. *Personality and Individual Differences*, *41*, 131–142.
- Brown KG, Le H, Schmidt FL. (2006). Specific Aptitude Theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment*, *14*, 87–100.
- Campbell JP, Gasser MB, Oswald FL. (1996). The substantive nature of job performance variability. In Murphy KR (Ed.), *Individual differences and behavior in organizations* (pp. 258–299). San Francisco: Jossey-Bass.
- Chernyshenko OS, Stark S, Prewett MS, Gray AA, Stilson FRB, Tuttle MD. (2006, May). Normative score comparisons from single stimulus, unidimensional forced choice, and multidimensional forced choice personality measures using item response theory. In Chernyshenko OS (Chair), *Innovative response formats in personality assessments: Psychometric and validity investigations*. Symposium conducted at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, Texas.
- Christiansen ND, Burns GN, Montgomery GE. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267–307.
- Christiansen ND, Goffin RD, Johnston NG, Rothstein MG. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *PERSONNEL PSYCHOLOGY*, *47*, 847–860.
- Colquitt JA, LePine JA, Noe RA. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85*, 678–707.
- Connelly BS, Ones DS. (2007a, April). *Combining Conscientiousness scales: Can't get enough of the trait, baby*. Poster session presented at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York City, NY.
- Connelly BS, Ones DS. (2007b, April). Multiple measures of a single Conscientiousness trait: Validities beyond .35! In Ones DS (Chair). *Too much, too little, too unstable: Optimizing personality measure usefulness*. Symposium conducted at the 22nd

- Annual Conference of the Society for Industrial and Organizational Psychology, New York City, NY.
- Cooper HM, Rosenthal R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, *87*, 442–449.
- Crant JM. (1995). The Proactive Personality Scale and objective job performance among real estate agents. *Journal of Applied Psychology*, *80*, 532–537.
- Dicken C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Educational and Psychological Measurement*, *23*, 699–720.
- Dilchert S. (2007). Peaks and valleys: Predicting interests in leadership and managerial positions from personality profiles. *International Journal of Selection and Assessment*, *15*, 317–334.
- Dilchert S, Ones DS, Viswesvaran C, Deller J. (2006). Response distortion in personality measurement: Born to deceive, yet capable of providing valid self-assessments? *Psychology Science*, *48*, 209–225.
- Dudley NM, Orvis KA, Lebiecki JE, Cortina JM. (2006). A meta-analytic investigation of Conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, *91*, 40–57.
- Dunlap WP, Cornwell JM. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, *29*, 115–126.
- Dunnette MD, McCartney J, Carlson HC, Kirchner WK. (1962). A study of faking behavior on a forced-choice self-description checklist. *PERSONNEL PSYCHOLOGY*, *15*, 13–24.
- Ellingson JE, Sackett PR, Connelly BS. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology*, *92*, 386–395.
- Ellingson JE, Sackett PR, Hough LM. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology*, *84*, 155–166.
- Erez A, Judge TA. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology*, *86*, 1270–1279.
- Foldes HJ, Ones DS, Sinangil HK. (2006). Neither here, nor there: Impression management does not predict expatriate adjustment and job performance. *Psychology Science*, *48*, 357–368.
- Frei RL, McDaniel MA. (1998). Validity of customer service measures in personnel selection: A review of criterion and construct evidence. *Human Performance*, *11*, 1–27.
- Gaugler BB, Rosenthal DB, Thornton GC, Bentson C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, *72*, 493–511.
- Goffin RD, Christiansen ND. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment*, *11*, 340–344.
- Gough HG. (1984). A managerial potential scale for the California Psychological Inventory. *Journal of Applied Psychology*, *69*, 233–240.
- Gough HG. (2005, January). Address given at the annual meeting of the Association for Research in Personality, New Orleans, LA.
- Guion RM, Gottier RF. (1965). Validity of personality measures in personnel selection. *PERSONNEL PSYCHOLOGY*, *18*, 135–164.
- Hedges LV, Olkin I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, *88*, 359–369.
- Heggestad ED, Morrison M, Reeve CL, McCloy RA. (2006). Forced-choice assessments of personality for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology*, *91*, 9–24.

- Hicks LE. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin*, *74*, 167–184.
- Hogan J, Holland B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*, 100–112.
- Hough LM. (1992). The “Big Five” personality variables—construct confusion: Description versus prediction. *Human Performance*, *5*, 139–155.
- Hough LM. (1998a). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance*, *11*, 209–244.
- Hough LM. (1998b). Personality at work: Issues and evidence. In Hakel MD (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection*. (pp. 131–166). Mahwah, NJ: Erlbaum.
- Hough LM, Ones DS. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In Anderson N, Ones DS, Sinangil Kepir H, Viswesvaran C (Eds.), *Handbook of industrial, work, and organizational psychology* (Vol. 1: Personnel psychology, pp. 233–277). London: Sage.
- Hough LM, Ones DS, Viswesvaran C. (1998, April). Personality correlates of managerial performance constructs. In Page R (Chair). *Personality determinants of managerial potential, performance, progression and ascendancy*. Symposium conducted at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Hough LM, Oswald FL. (2000). Personnel selection: Looking toward the future—Remembering the past. *Annual Review of Psychology*, *51*, 631–664.
- Hunter JE, Schmidt FL. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Hunter JE, Schmidt FL. (2004). *Methods of meta-analysis*. Thousand Oaks, CA: Sage.
- Hurtz GM, Donovan JJ. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, *85*, 869–879.
- Ironson GH, Davis GA. (1979). Faking high or low creativity scores on the Adjective Check List. *Journal of Creative Behavior*, *13*, 139–145.
- James LR, McIntyre MD, Glisson CA, Green PD, Patton TW, LeBreton JM, et al. (2005). A conditional reasoning measure for aggression. *Organizational Research Methods*, *8*, 69–99.
- Johnson CE, Wood R, Blinkhorn SF. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, *61*, 153–162.
- Judge TA, Bono JE, Ilies R, Gerhardt MW. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*, 765–780.
- Judge TA, Cable DM, Boudreau JW, Bretz RD. (1995). An empirical investigation of the predictors of executive career success. *PERSONNEL PSYCHOLOGY*, *48*, 485–519.
- Judge TA, Erez A. (2007). Interaction and intersection: The constellation of emotional stability and extraversion in predicting performance. *PERSONNEL PSYCHOLOGY*, *40*, 573–596.
- Judge TA, Erez A, Bono JE, Thoresen CJ. (2003). The Core Self-Evaluations Scale: Development of a measure. *PERSONNEL PSYCHOLOGY*, *56*, 303–331.
- Judge TA, Heller D, Mount MK. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, *87*, 530–541.
- Judge TA, Higgins CA, Thoresen CJ, Barrick MR. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *PERSONNEL PSYCHOLOGY*, *52*, 621–652.
- Judge TA, Ilies R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, *87*, 797–807.

- Le H, Ilies R, Holland EV. (2007, April). Too much of a good thing? Curvilinearity between Emotional Stability and performance. In Ones DS (Chair). *Too much, too little, too unstable: Optimizing personality measure usefulness*. Symposium conducted at the 22nd Annual Conference of the Society for Industrial and Organizational Psychology, New York City, New York.
- LePine JA, Erez A, Johnson DE (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, 87, 52–65.
- Li A, Bagger J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment*, 14, 141.
- 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.
- Marshall MB, De Fruyt F, Rolland J-P, Bagby RM. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. *Psychological Assessment*, 17, 379–384.
- McCloy RA, Heggstad ED, Reeve CL. (2005). A silk purse from the sow's ear: Retrieving normative information from multidimensional forced-choice items. *Organizational Research Methods*, 8, 222–248.
- McDaniel MA, Hartman NS, Whetzel DL, Grubb W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *PERSONNEL PSYCHOLOGY*, 60, 63–91.
- McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- McDaniel MA, Whetzel DL, Schmidt FL, Maurer SD. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- Meade AW. (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, 531–552.
- Mol ST, Born MPH, Willemsen ME, Van Der Molen HT. (2005). Predicting expatriate job performance for selection purposes: A quantitative review. *Journal of Cross-Cultural Psychology*, 36, 620.
- Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *PERSONNEL PSYCHOLOGY*, 60, 683–729.
- Mount MK, Barrick MR, Strauss J. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology*, 79, 272–280.
- Murphy KR, Shirella AH. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *PERSONNEL PSYCHOLOGY*, 50, 823–854.
- Ng TWH, Eby LT, Sorensen KL, Feldman DC. (2005). Predictors of objective and subjective career success. A meta-analysis. *PERSONNEL PSYCHOLOGY*, 58, 367–408.
- Nunnally JC. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Ones DS. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation, University of Iowa, Iowa City, IA.
- Ones DS, Hough LM, Viswesvaran C. (1998, April). *Validity and adverse impact of personality-based managerial potential scales*. Poster session presented at the 13th

- Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Ones DS, Viswesvaran C. (2001a). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9, 31–39.
- Ones DS, Viswesvaran C. (2001b). Personality at work: Criterion-focused occupational personality scales used in personnel selection. In Roberts BW, Hogan R (Eds.), *Personality psychology in the workplace* (pp. 63–92). Washington, DC: American Psychological Association.
- Ones DS, Viswesvaran C. (in press). Customer service scales: Criterion-related, construct, and incremental validity evidence. In Deller J (Ed.), *Research contributions to personality at work*.
- Ones DS, Viswesvaran C, Dilchert S. (2004). Cognitive ability in selection decisions. In Wilhelm O, Engle RW (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431–468). Thousand Oaks, CA: Sage.
- Ones DS, Viswesvaran C, Dilchert S. (2005a). Cognitive ability in personnel selection decisions. In Evers A, Voskuijl O, Anderson N (Eds.), *Handbook of selection* (pp. 143–173). Oxford, UK: Blackwell.
- Ones DS, Viswesvaran C, Dilchert S. (2005b). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, 18, 389–404.
- Ones DS, Viswesvaran C, Reiss AD. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679.
- Ones DS, Viswesvaran C, Schmidt FL. (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.
- Organ DW, Ryan K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *PERSONNEL PSYCHOLOGY*, 48, 775–802.
- Peeters MAG, Van Tuijl HFJM, Rutte CG, Reymen IMMJ. (2006). Personality and team performance: A meta-analysis. *European Journal of Personality*, 20, 377–396.
- Ree MJ, Earles JA, Teachout MS. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, 79, 518–524.
- Robie C, Zickar MJ, Schmit MJ. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance*, 14, 187–207.
- Rosse JG, Stecher MD, Miller JL, Levin RA. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634–644.
- Rothstein HR, Schmidt FL, Erwin FW, Owens WA, Sparks PC. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology*, 75, 175–184.
- Ruch FL, Ruch WW. (1967). The k factor as a (validity) suppressor variable in predicting success in selling. *Journal of Applied Psychology*, 51, 201–204.
- Rust J. (1999). The validity of the Giotto integrity test. *Personality and Individual Differences*, 27, 755–768.
- Salgado JF. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30–43.
- Schmidt FL, Hunter JE. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.

- Schmit MJ, Ryan AM, Stierwalt SL, Powell AB. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–620.
- Schmitt N, Oswald FL. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 91*, 613–621.
- Schmitt N, Rogers W, Chan D, Sheppard L, Jennings D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719–730.
- Smith DB, Ellingson JE. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87*, 211–219.
- Smith DB, Hanges PJ, Dickson MW. (2001). Personnel selection and the five-factor model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology, 86*, 304–315.
- Stark S, Chernyshenko OS, Chan K-Y, Lee WC, Drasgow F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86*, 943–953.
- Stark S, Chernyshenko OS, Drasgow F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-dimensional pairwise-preference model. *Applied Psychological Measurement, 29*, 184–203.
- Stark S, Chernyshenko OS, Drasgow F, Williams BA. (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.
- Steel P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin, 133*, 65–94.
- Tenopyr ML. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology, 73*, 749–751.
- Topping GD, O'Gorman, J. (1997). Effects of faking set on validity of the NEO-FFI. *Personality and Individual Differences, 23*, 117–124.
- Vasilopoulos NL, Cucina JM, Dyomina NV, Morewitz CL, Reilly RR. (2006). Forced-choice personality tests: A measure of personality and cognitive ability? *Human Performance, 19*, 175–199.
- Villanova P, Bernardin HJ, Johnson DL, Dahmus SA. (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 DS. (2000). Measurement error in "Big Five factors" personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement, 60*, 224–235.
- Viswesvaran C, Ones DS, Hough LM. (2001). Do impression management scales in personality inventories predict managerial job performance ratings? *International Journal of Selection and Assessment, 9*, 277–289.
- Witt LA, Burke LA, Barrick MA, Mount MK. (2002). The interactive effects of Conscientiousness and agreeableness on job performance. *Journal of Applied Psychology, 87*, 164–169.
- Zhao H, Seibert SE. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology, 91*, 259–271.

APPENDIX
Studies Listed by Campion as Indicating That Response Distortion Affects Criterion-Related Validity of Personality Measures

	Directed faking/ lab study with assignment	Studied nonmotivating contexts	Studied job applicants/ motivated test takers
Did not study criterion-related validity	◆ Abbott and Harris (1973) (80) [<i>Student sample; did not include any criterion measures.</i>]		◆ Rosse, Stecher, Miller, and Levin (1998) (11) [<i>Did not include any criterion measures.</i>]
Did not include performance criterion	◆ Topping and O'Gorman (1997) (12) [<i>Student sample; criterion was peer-rating of personality.</i>] ◆ Ironson and Davis (1979) (64) [<i>Student sample, criterion was rating of creativity.</i>]		
Primary study of criterion related validity	◆ Schmit, Ryan, Stierwalt, and Powell (1995) (19) [<i>Mixed findings: validity was affected when students were responding as applicants, but when the personality measure was contextualized, validity was strong and higher than for the nonapplicant condition.</i>]	◆ Dicken (1963) (98) [<i>Mostly high-school and college students; correcting for distortion did NOT result in increases in validity beyond chance level.</i>]	◆ Ruch and Ruch (1967) (93) [<i>Results in opposite direction: Sizeable criterion-related validity existed, but dropped when corrected for purported faking.</i>]
	◆ Dunnette et al. (1962) (101) [<i>Validity among 45 incumbents instructed to fake was investigated.</i>]		

Note. Numbers in parentheses correspond to those listed in the reference list of Morgeson et al. (2007).