Personality and Cognitive Ability as Predictors of Effective Performance at Work

Neal Schmitt

Department of Psychology, Michigan State University, East Lansing, Michigan 48824;
email: schmitt@msu.edu

Keywords

selection, test validity, general factor, diversity, faking, cheating, multilevel issues

Abstract

Conclusions about the validity of cognitive ability and personality measures based on meta-analyses published mostly in the past decade are reviewed at the beginning of this article. Research on major issues in selection that affect the use and interpretation of validation data are then discussed. These major issues include the dimensionality of personality, the nature and magnitude of *g* in cognitive ability measures, conceptualizations of validity, the nature of the job performance domain, trade-offs between diversity and validity, reactions to selection procedures, faking on personality measures, mediator and moderator research on test–performance relationships, multilevel issues, Web-based testing, the situational framing of test stimuli, and the context in which selection occurs.

INTRODUCTION

Building on impressive recent work in the area of personality and cognitive ability as predictors of performance, this review focuses on some major issues regarding the use, implications, and refinement of measures of these variables. Indeed, it would seem unnecessary to review the basics in this area, as multiple excellent recent reviews are available. Sackett & Lievens (2008) provided the last updating of this literature in the *Annual Review of Psychology*. Ones et al. (2012) provided a meta-analytic review of general cognitive ability as well as specific aspects of cognitive ability as predictors of training success and performance. A more general narrative review of the validity of cognitive ability tests was provided by Ones et al. (2010). Schmitt & Fandre (2008) presented a summary of various meta-analyses of general cognitive ability as well as specific mental abilities. Hulsheger et al. (2007) provided an analysis of the validity of cognitive ability tests in Germany, and Salgado and colleagues (Salgado et al. 2003a,b) reported similar investigations in the European community. Two comprehensive handbooks on selection (Farr & Tippins 2010, Schmitt 2012) have also appeared in the past several years.

Data regarding the validity of personality measures are also widely available and have been meta-analyzed frequently. Beginning with Barrick & Mount (1991), there have been numerous meta-analyses that establish a relationship between some aspects of the so-called five-factor model of personality (the Big Five constructs or personality dimensions, or simply the Big Five) and job performance. Barrick et al. (2001) provided a meta-analysis of meta-analyses of what was at that time a burgeoning effort. More recently, Barrick & Mount (2012) provided a review of the relationship between personality and various work outcomes, and Hough & Dilchert (2010, p. 309) provided a summary of the relationships between various personality constructs and aspects of work performance.

The results of these reviews and others lead to the following conclusions:

1. The validity of cognitive ability is generalizable across situations; the observed correlation between job performance and measured cognitive ability is usually in the .20s, and the validity corrected for range restriction and/or criterion unreliability is most often .40 or above.
2. The relationships between personality measures and performance vary with which of the Big Five constructs one considers and appear to be generalizable only in the case of conscientiousness. Observed correlations between performance and individual measures of personality are almost always less than .20, and corrected correlations rarely exceed .25. A spirited exchange regarding the validity and use of personality tests in selection was provided by Morgeson et al. (2007), Ones et al. (2007), and Tett & Christiansen (2007).
3. Correlations between cognitive ability measures and personality measures are usually low. Judge et al. (2007) reported that meta-analytic correlations between four of the Big Five constructs and cognitive ability were less than .10; openness correlated .22 with cognitive ability. Similarly, Roth et al. (2011) estimated the corrected correlation between conscientiousness and cognitive ability to be .03.
4. Correlations between measures of the Big Five constructs are usually moderate (less than .40). Van der Linden et al. (2010) reported meta-analytic correlations, in absolute terms based on 212 separate correlation matrices and the responses of 144,117 individuals, that ranged between .12 and .32 (.17 and .43, respectively, when corrected for unreliability).
5. These statements in combination suggest that both cognitive ability and personality are valuable predictors of job performance and, given their relative lack of correlation with each other, that combinations of the two will produce superior predictions of job performance. In addition, the low intercorrelations of Big Five measures suggest that

personality combinations are also likely to produce validities that are larger in magnitude than are the validities of the individual Big Five constructs. Further discussion of both of these conclusions follows below.

As noted above, and given that the basic questions regarding the validity of cognitive ability and personality have been addressed and the results as summarized above are widely accepted, the focus of this review is on some major issues regarding the use, implications, and refinement of measures of these variables. Addressing these issues constitutes the thrust of the majority of the literature in this area over the past decade. Specifically, the following theoretical and practical questions/issues are addressed in this review.

1. Are the Big Five an adequate explanatory taxonomy for personality? Most researchers have accepted this structure as a reasonable representation of the dimensions of human personality, and this structure has served to organize a very diverse literature. However, Hough & Oswald (2005) among others have argued consistently that the Big Five measures are too broad to represent the personality determinants of well-developed and specific performance criteria. However, there are a group of scholars in the personality area who believe there are one or two general dimensions of personality that underlie more specific measures, including measures of the Big Five.

2. Even though a general factor (g) appears to account for the majority of the variance in ability measures, some research on the nature and dimensionality of intelligence suggests there are clear differences between crystallized and fluid intelligence (Nisbett et al. 2012), and there continues to be exploration of the importance of various specific abilities (in particular, perceptual speed and accuracy) and their use in combination with general mental ability (Campbell & Catano 2004, Mount et al. 2008).

3. How are tests validated, and what does validity mean? The types of evidence used to support claims of validity continue to evolve, as evidenced by versions of the APA Standards, but new ideas are also being presented in our scientific literature.

4. The field also seems to be considering multiple performance domains as opposed to overall task performance, and the number of performance domains seems to be expanding (e.g., Chiaburu et al. 2011). This expansion obviously has implications for the utility of different types of predictors.

5. Models of the trade-offs between diversity and validity (DeCorte et al. 2010) continue to be explored to provide insight into the impact of test use on outcomes that are often in conflict. How predictors are combined and weighted influences this trade-off. In addition, researchers have continued to compile data regarding subgroup differences in validity (Berry et al. 2011).

6. Are various tests perceived to be fair, and how does this impact what is measured and how the information about applicants is used (Gilliland & Steiner 2012)?

7. With increased use of personality measures, questions regarding the role of faking and other response sets continue to concern test users.

8. Hypotheses about predictor–criterion relationships have become more sophisticated and include a consideration of moderator and mediator effects involving multiple personality variables as well as personality and cognitive ability.

9. There has been considerable discussion (e.g., Ployhart & Moliterno 2011) and some empirical research about the nature of predictor–criterion relationships at different levels (group, team, organization, etc.) of analysis.

10. Test items, especially personality-type items, appear to be more acceptable and valid when they are framed to appear face valid or context specific (Shaffer & Postlethwaite 2012).

11. Several issues are of a more practical than theoretical nature. The advent of Web-based testing has raised a number of issues about proctoring test takers and the quality of data collected in unsupervised settings, as well as how to examine and control quality concerns and test security.

12. Finally, a host of issues related to the organizational or societal context in which selection occurs have occupied the attention of a growing body of practitioners and researchers (Tippins 2012).

PERSONALITY DIMENSIONALITY

Following Barrick & Mount (1991), the vast majority of investigations of the personality correlates of performance have used the Big Five taxonomy as the basis of their selection of predictors. However, there is an increasing body of evidence supporting Hough's contention that better predictions of performance can be achieved by using measures more narrowly tailored to the particular prediction task at hand (Hough 1992, Hough & Ones 2001, Hough & Oswald 2005).

Consistent with this notion, Bergner et al. (2010) found that narrow traits added incrementally beyond the Big Five to the prediction of managerial success (salary progression and supervisory ratings). Perry et al. (2010) found that only one (i.e., achievement) of six conscientiousness facets interacted with general mental ability to impact the performance of customer service representatives. Van Iddekinge et al. (2005) found that two facets of a broad integrity test were correlated more highly with job performance than was the broad integrity scale and that the multiple correlation of these two facets was more than three times as large as the validity of the broad integrity test.

Similar results have been found when predicting other aspects of performance. For example, workplace deviance was not predicted by neuroticism but was predicted by anger, one of its facets. Moreover, five facets picked by experts a priori as being related to deviance predicted nearly as much variance as did the Big Five in the reported deviance of participants in a psychology class who had work experiences (Hastings & O'Neill 2009). Casillas et al. (2009) found that a risk reduction facet of an integrity measure provided incremental validity over general work attitudes in predicting safety behavior and counterproductive work behavior as reflected in supervisory ratings.

The studies summarized above all involve consideration of facets of the Big Five. Hough & Schneider (1996) proposed that the Big Five model (and its facets) provides a reasonable taxonomic summary of personality measures but that other variables such as achievement, social insight, affiliation, and rugged individualism merit further attention as basic personality constructs, especially when we are interested in predicting work performance.

Another important and related development is the recognition that compound personality variables (combinations of basic personality variables), such as integrity, customer service orientation (Frei & McDaniel 1998), and core self-evaluation (Judge & Bono 2001), often provide better predictions of outcome variables than do single aspects of the Big Five model. In addition, weighted composites of the Big Five variables or other personality variables, given their minimal intercorrelation and their relevance to a desired outcome, are likely to produce multiple correlations that are substantially higher than the validities reported in meta-analyses that typically report data for each construct individually. However, if we are to produce generalizable personality research, it seems best not to produce a proliferation of compound personality variables without careful retention of their nature vis-à-vis the Big Five structure including its facets.

Going in quite a different direction, some basic personality researchers argue that there is a general factor of personality, with two or more factors lying beneath the general factor in a hierarchical model (Van der Linden et al. 2010). However, other personality researchers (e.g., Hopwood et al. 2011) have not found evidence for the general factor in analyses of responses to several personality inventories. Given the typical correlations between measures of the Big Five (see above), it is certainly the case that higher-order factors of personality, if they exist in quantifiable form, are unlikely to account for a great deal of variance, certainly nothing like g in ability testing.

For practitioners using personality measures, it seems that it would be best to thoughtfully choose facets of personality relevant to a specific criterion in a given situation a priori. In combination, these facets are likely to produce the highest criterion-related validities (e.g., Paunonen & Ashton 2001). Such rationality in the choice and validation of predictors is also likely to be of most value scientifically. For similar arguments as to why personality validities are not of greater magnitude (and by implication, how they can be improved), see Murphy & Dzieweczynski (2005).

IS ABILITY A GENERAL FACTOR?

As indicated above, a general factor, or g, appears to be present in most ability tests; g is an effective predictor of job performance, and arguably the most important predictor of performance (Reeve & Hakel 2002). Cognitive abilities are usually represented in a hierarchy, with g at the top, followed by relatively broad content categories such as verbal and numerical ability, followed by narrower, more specific abilities such as spelling or word knowledge (Carroll 1993). In their studies of the training success and performance of military personnel, Ree and colleagues reported that most of the validity of cognitive tests typically comes from g and that measures of specific ability increment the validity of g-based measures little, if at all (e.g., Carretta & Ree 2000, Ree & Carretta 2002). However, the magnitude of g and the degree to which measures of specific abilities add incremental validity to measures of g continue to be investigated. Campbell & Catano (2004) reported that auditory attention improved the prediction of training performance among Canadian military personnel in the "Operator Family." Mount et al. (2008) reported that a test of perceptual speed and accuracy improved the prediction of task performance, above general mental ability, in a group of warehouse workers. In the latter study, Mount et al. made the point that the importance of g for the performance of these relatively low-level jobs was likely low, leaving room for the predictive efficacy of other abilities.

Using a meta-analytic database developed using responses to a German cognitive ability test based on Thurstone's (1938) primary mental abilities, Lang et al. (2010) found, when traditional multiple regression analyses were used, that g accounted for about 80% of the criterion-related variance and that various primary mental abilities accounted for 20% of the criterion-related variance. When relative importance analyses (Johnson & LeBreton 2004) were used to analyze the data, the contribution of verbal comprehension exceeded that of g, and several other primary abilities provided contributions similar to that of g. In addition, Nisbett et al. (2012) argued that crystallized intelligence (derived from life experience and education) and fluid intelligence (native ability) are "quite different aspects of intelligence at both the behavioral and biological levels" (p. 130). They also speculated that g could derive from largely independent cognitive skills that coalesce over time and educational experience.

Although it is important to continue research into the nature of g and other specific abilities, it is probably wisest to continue to select using g when possible. Given that g appears to be

the ability to learn and process information, which are important to the acquisition of knowledge and skill, it would appear to be widely relevant in all jobs. A large body of meta-analytic research is consistent with the recommendation that g be used in selection for most, if not all, jobs (e.g., Schmidt & Hunter 1998).

EVIDENCE OF TEST VALIDITY

In assessing the use of personality and cognitive ability in the selection context, we try to support the inference that scores on tests of these traits can be used to draw inferences about future performance. The evidence used to support this inference is termed validity. The way that scientists have conceptualized validity has changed over the past several decades, as documented in the several versions of the APA Standards (AERA et al. 1999) and SIOP (2003) Principles. The most recent versions of both these documents treat validity as a unitary concept that is supported by a variety of evidence. However, thinking about validity continues to develop, as evidenced in two excellent reviews of the literature (see Kehoe & Murphy 2010, Sackett et al. 2012), a book on alternative strategies of validation (see McPhail 2007), and spirited exchanges about the nature of content evidence for validity and about synthetic validity (see issue 4 of the 2009 volume and issue 3 of the 2010 volume of *Industrial and Organizational Psychology: Perspectives on Science and Practice*, respectively).

Van Iddekinge & Ployhart (2008) summarized the implications of several new developments to the estimation of criterion-related validity. They pointed out that underestimates of validity are likely if one does not correct for artifacts, but the correct formulas and estimates of the unrestricted range are often difficult to ascertain. Estimates of indirect range restriction are likely more accurate estimates of the impact of range restriction (see Schmidt et al. 2006). Sackett & Yang (2000) described 11 different cases of range restriction that can influence the use of correction formula and statistics.

In evaluating multiple predictors, a small body of research has promoted the use of relative weight analyses (LeBreton et al. 2004). Sackett et al. (2003) underscored the importance, when addressing the bias in the target variable using regression analyses, of including omitted variables that are correlated with both the criterion and the target predictor. Not doing so is likely to suggest bias in the target variable when the bias is a function of some other omitted variable. With respect to using criteria such as organizational citizenship behavior, adaptive behavior, and counter-productive work behavior, Van Iddekinge & Ployhart (2008) recommended caution. These alternate forms of work behavior often do not appear in a traditional job analysis oriented to task performance (necessitating some alternate means of justification for their use), and they may be highly correlated with task performance (see also the section below on criteria). Van Iddekinge and Ployhart did report a few studies on maximum and typical performance; most studies have found that these performance indices are not highly correlated and that they do have different correlates (e.g., Marcus et al. 2007).

Johnson & Carter (2010) provided an excellent case study of the application of synthetic validity. Using job analysis data, the authors identified 11 job families and 27 job components. They created a test composite for each job component and then, based on relevant job components, chose a test battery for each job family. Next, they computed synthetic validity coefficients for each test battery and compared the coefficients with traditional validity coefficients computed for the large job families. The synthetic validity coefficients were very similar to traditional criterion-related validities. In another paper, Johnson et al. (2010) described two types of synthetic validation and argued that researchers should develop a comprehensive database to create prediction equations for use in synthetic validation of jobs across the US economy.

Responses to this proposal are contained in issue 3 in the 2010 volume of *Industrial and Organizational Psychology: Perspectives on Science and Practice.*

CONTINUED EXPANSION OF THE PERFORMANCE DOMAIN

For well over a half century, personnel selection researchers routinely examined the relationship between cognitive ability, personality, and job performance using a supervisory rating of overall job performance or a sum of several highly correlated ratings of task performance. Occasionally, predictors were related to other outcomes, such as turnover, absenteeism, and accidents.

Campbell et al. (1993) provided a broader conceptualization of job performance, and since then, researchers have considered an increasingly large number of performance outcomes. Borman & Motowidlo (1993) introduced the notion that employees often add to workplace effectiveness in ways that are not directly related to the tasks specified in their job descriptions. These contextual behaviors included volunteering to do special tasks not part of one's job description, helping and cooperating with others, and supporting organizational objectives. A related concept is referred to as organizational citizenship behavior (Organ 1997).

In the area of poor performance, researchers have focused on counterproductive work behavior such as theft, aggression, violence, destruction of property, and withdrawal (Rotundo & Spector 2010), and some have maintained that counterproductive work behavior itself is multidimensional (see Berry et al. 2007 for a meta-analytic review). Hoffman & Dilchert (2012) provided a recent review of the personality correlates of organizational citizenship behavior (see Chiaburu et al. 2011 for a meta-analysis) and counterproductive work behavior. Integrity tests (often a composite of several of the Big Five constructs, especially conscientiousness) have long been used to predict counterproductive work behavior, albeit with some degree of controversy regarding their levels of validity (see Van Iddekinge et al. 2012 and related papers in the same issue).

There has also been a burgeoning of research on adaptive performance and its predictability (Pulakos et al. 2012). More traditional outcomes such as turnover and safety continue to be of interest to organizations and researchers, but they are often seen as part of a more general withdrawal construct (Harrison & Newman 2012).

This expansion of the performance domain has in turn expanded the set of predictors that are of value in a personnel selection context, and this has been particularly true for personality constructs. To be useful, however, these performance constructs must not be highly redundant with either measures of task performance or each other. This high redundancy seems to be true for the distinction between organizational citizenship behavior (or contextual performance) and task performance; it may be less true for adaptive performance and task performance (Pulakos et al. 2002). Counterproductive work behavior and organizational citizenship behavior are not simply opposite ends of a single performance continuum (Dalal 2005). Withdrawal measures are usually minimally correlated with task performance. This literature on the nature and dimensionality of performance highlights the necessity of considering carefully which aspects of performance are relevant in a particular organizational context and that predictors be selected accordingly. It also suggests that we begin to investigate multiattribute models of performance, as Murphy & Shiarella (1997) have suggested.

DIVERSITY-VALIDITY DILEMMA AND PREDICTIVE BIAS

For at least the past couple of decades, researchers found that the use of tests did not produce biased predictions for women and members of minority groups. The usual conclusion was that

there was some small overprediction of minority performance based on intercept differences found in regressions of performance outcomes on test scores, minority status, and their interaction. Both this method of analysis and the empirical research have been institutionalized in guidelines (see AERA et al. 1999, SIOP 2003). Recently, however, there have been challenges to the notion that minority outcomes are underpredicted by psychological tests. Aguinis et al. (2010) have argued that test bias research should be revived. In a simulation, they showed that nearly all previous tests of slope bias had insufficient power to detect a difference in slopes. They also argued that the test of intercept differences itself is biased when minority scores are lower than majority scores and when the test has relatively low reliability. Finally, they proposed that future test bias research consider time and context of selection as possible determinants of slope and intercept differences between groups.

Berry et al. (2011) provided a meta-analysis of cognitive ability validity coefficients across black, white, Hispanic, and Asian subgroups. They found white and Asian validities to be .33, whereas average black validity (.24) and Hispanic validity (.30) were lower. Black–white differences were especially pronounced in the military context, as opposed to civilian and educational contexts. Asian and Hispanic differences were available only for educational contexts. Like Aguinis et al. (2010), Berry et al. (2011) called for additional research on differential validity, which contradicts earlier strong statements that such differences do not exist (e.g., Hunter et al. 1979).

Mattern & Patterson (2013) have provided such reexamination of differential prediction using the SAT to predict college student performance. Data from 177 institutions and more than 450,000 students were used to assess differential prediction of first-year grade point average. The authors found minimal differential prediction, a slight overprediction of the grades of black and Hispanic students, and slight underprediction of female performance, all of which corroborate early findings in this area. No similar analysis of employment data is available at this time. Mattern and Patterson also provided each institution's data in a supplementary file, which should facilitate future analyses. If institutional context variables are available, the type of work described in Aguinis et al. (2010) should be possible. Additional research on the issues raised by Aguinis et al. will certainly appear.

Another issue that continues to receive attention is the conflict between the goals of maximizing organizational effectiveness and a diverse workforce when using a valid test (or test battery) that displays mean differences in subgroup performance. In a series of papers, DeCorte and colleagues (DeCorte et al. 2007, 2010, 2011; Sackett et al. 2008) provided an analytic technique that optimizes the goals prescribed by an organization. The various aspects that are entered into the optimization procedure include the predictor subset and its characteristics, the selection ratio or selection rule, the staging and sequencing of the selection procedure, and the weighting of the predictors. They showed that there is a great range of optimal solutions given these inputs. This approach can help decision makers determine the type of selection system they want to design given the particular context in which they function. There are, of course, other factors that influence whether an organization can meet diversity and productivity goals. Newman & Lyon (2009) described the constraints of the recruitment process on the realization of diversity goals, and Tam et al. (2004) examined the impact of applicant withdrawal on desired system outcomes. Although this research is mostly simulation or Monte Carlo work, it has added to our understanding of the practical constraints under which selection researchers and practitioners must work when designing selection systems that must serve sometimes conflicting goals. How tests are weighted and used in this situation determines the degree to which an organization can meet its goals (see Hattrup 2012 for a review of the literature on weighting).

Finally, Ployhart & Holz (2008) provided a relatively nontechnical review of these and other strategies that can serve to balance diversity and validity. Additionally, Roth et al. (2011) have

provided estimates of the validity of different measures, their intercorrelations, and the level of subgroup differences for typical combinations, which can be used to maximize diversity and validity by combining measures that display minimal subgroup differences with those for which we usually find much larger subgroup differences (e.g., cognitive ability). These data are helpful for developing realistic expectations regarding the efficacy of this approach.

PERCEPTIONS OF SELECTION TESTS

Over the past two decades, researchers and practitioners have recognized that the examinee is an active part of an assessment and has perceptions of the event that affect future perceptions and behavior vis-à-vis the assessor or assessing organization. In fact, there is probably more research on this issue than on any other related to personnel selection in the past 20 years. Gilliland & Steiner (2012) provided a comprehensive review of this literature. Surprisingly, their review indicated that perceptions of fairness in selection are still correlated primarily with other perceptions or self-reports of behavior. Very little research has been conducted relating perceptions of the selection process or instruments to behavioral outcomes, and the few studies that have been done reported very small effect sizes. Research regarding job candidates' prehire decisions, such as electing to withdraw from the hiring process, indicates that such decisions do not seem to be related to the candidates' perceptions of the process itself (Ryan et al. 2000, Schmit & Ryan 1997).

In a meta-analysis comparing reactions to different types of tests, Anderson et al. (2010) found that reactions to personality and cognitive ability tests were evaluated favorably though work samples and interviews were most favorably viewed by applicants. That these perceptions are important is underscored in a paper by Rynes et al. (2002) in which the authors reported that practitioners' beliefs regarding the efficacy of cognitive ability and personality tests were much less than warranted by the research literature. A meta-analytic review by Truxillo et al. (2009) indicated that explanations of the nature of the selection process by company personnel or others affect applicants' fairness perceptions, perceptions of the employing organization, test-taking motivation, and performance on cognitive ability tests. As these explanations will most likely be provided by practitioners in the selection context, the results of the Rynes et al. study are relevant in that context as well.

The overall conclusion from this relatively large body of research is that fairness perceptions do influence applicants' reactions to the hiring process and the organization in which they hope to work, but that there is a paucity of research connecting these perceptions to behavior of these applicants either during the employment decision process or later on the job. However, perceptions of tests may affect the degree to which human resource specialists or other organizational decision makers use such instruments.

FAKING OF PERSONALITY TESTS/CHEATING ON ABILITY TESTS

As would be expected given the proliferation of Internet testing, there is increasing concern that examinees, particularly those in high-stakes situations, will fake "good" on personality measures and attempt to cheat on ability or knowledge tests. Although there remains some debate on the extent of faking on personality tests (Hogan et al. 2007), most researchers agree that personality measures can be faked and that some applicants do attempt to manage the impressions conveyed by their responses to personality measures (Birkeland et al. 2006). Responses to ability measures are different in that there are right and wrong answers to items, and dishonest examinees strive to get and record the right answer in a variety of ways (Arthur et al. 2009, Hausknecht et al. 2007). With a large item pool and a computer-adaptive test, it is very difficult, but certainly

not impossible, to cheat on an ability test. However, cheaters can often be relatively easily detected.

Accordingly, there seems to be more interest in how to evaluate and deter cheating, or response distortion, on personality measures. Perhaps the oldest method of correcting for presumed distortion on personality scales is to include response-distortion or lie scales and use the scores on these scales to either correct scores on substantive scales or discard the responses of some respondents (Goffin & Christiansen 2003). The most modern approach to the detection of faking may be represented by the use of eye trackers and response latency measures (Van Hooft & Born 2012). Reeder & Ryan (2011) have provided a review of these corrections for faking. However, the manner in which to use these corrections is often not clear, and recently, Schmitt & Oswald (2006) showed that with the usual level of personality test validity and the level of correlation with a response-distortion scale, even a perfect response-distortion scale would not change estimates of validity by any appreciable amount. Ellingson et al. (2012) reported an interesting study in which individuals suspected of cheating were asked to take a retest; the test takers' second scores appeared to be more accurate estimates of their standing on personality variables.

Aside from validity, however, there remains the problem that some individuals may get selected because of response distortion. In reviewing the literature on faking, Ziegler (2011) estimated that 30% of applicants fake. Certainly there are individual differences in the extent of faking. Estimates of these persons' job-related characteristics will be inflated relative to those of people who do not fake or who fake less. Several recent studies reported on attempts to reduce faking, including the use of warnings (e.g., Fan et al. 2012) and special questionnaire types (Stark et al. 2012). A comprehensive review of issues related to faking on personality measures is available in an edited book by Ziegler et al. (2011).

MEDIATOR AND MODERATOR EFFECTS IN PERSONALITY RESEARCH

Most validation research has focused on the linear relationship between predictor and criterion, and indeed a linear relationship appears to be relatively universal in ability–performance relationships (Coward & Sackett 1990). However, there are a number of recent studies that indicate that personality–performance relationships may be more complex and that consideration of interactive and mediating effects may increase our understanding of the role of individual differences in performance and, in some cases, may increase the predictive value of these variables. Among the first to investigate such relationships, Barrick & Mount (1993) found that the relationships between conscientiousness and extraversion and performance were stronger for managers whose jobs allowed a high degree of autonomy than for those with jobs with low autonomy, and Barrick et al. (1993) found that goal setting mediated the effects of conscientiousness on performance. More recently, Barrick et al. (2013) provided a theory of the interaction of personality traits, individual goals, and situational characteristics on performance.

Penney et al. (2011) provided a review of situational moderators of the personality–performance relationship and argued that it is likely that task, social, and organizational variables moderate this relationship. They also concluded that traits likely interact to impact performance. In another meta-analytic study, Meyer et al (2009) found that situational strength (constraints on performance and consequences of performance as coded from O*NET data on occupations involved in the primary studies) moderated the conscientiousness–performance relationship in that the relationship was greater when situational strength was high than when it was low in magnitude.

Witt (2002) found that extraversion led to greater levels of performance among highly conscientious workers but lower performance among those low in conscientiousness. Postlethwaite

et al. (2009) found that conscientiousness was a stronger predictor of safety behavior for individuals high in cognitive ability than for those with low levels of cognitive ability. Contradicting these findings, Sackett et al. (1998), using data from four very large databases, failed to find interactions between various personality characteristics and cognitive ability in predicting performance. Finally, interactions between personality variables and performance across time periods have also been reported (e.g., Thoresen et al. 2004, Zyphur et al. 2008).

Whetzel et al. (2010), using data from 112 financial service employees, found little evidence of any curvilinear relationships between performance and 32 scales of the Occupational Personality Questionnaire. However, Le at al. (2011) reported evidence for nonlinearity. Nonlinear relationships between conscientiousness and emotional stability and task performance of public service employees in a variety of jobs indicated that at high (versus low) levels of the trait, more of the trait was associated with decrements in performance. Moreover, this curvilinearity was moderated by the complexity of the job. Emotional stability was also curvilinearly related to organizational citizenship behavior.

In summary, personality–performance relationships may be complex, as some of these studies indicate. It also appears to be the case that such complex relationships are observed when they are preceded by careful theorizing as to why some interaction or mediation is likely. This theorizing would seem to be important, as the potential to find a few significant or sizable interactions that prove nonreplicable or spurious among the many possible is often quite high. Also, it should be noted that the relatively low reliability of some personality measures may obscure findings of curvilinearity.

MULTILEVEL ISSUES

Selection research grew out of an interest in individual differences, so it is natural that the primary focus has been on the nature of relationships among variables measured at the level of the individual. Particularly over the past couple of decades, however, researchers have discovered that relationships and processes are often different when we consider the behavior of individuals, groups of individuals, teams, or organizations and other aggregated units. Klein & Kozlowski (2000) introduced multilevel issues and applications to the organizational psychology community, but until recently, most treatments of multilevel issues in selection have been theoretical (e.g., Ployhart 2004, 2012; Ployhart & Moliterno 2011). In one exception, Ployhart et al. (2011) provided evidence that generic human capital in the form of personality and cognitive ability leads to changes in unit-specific human capital such as training and experience, which in turn lead to unit service performance and valued outcomes. In a similar study, Van Iddekinge et al. (2009) showed that the implementation of selection and training programs was related to customer service and training effectiveness, which in turn were related to restaurant profits. Selection in that study was operationalized as the percentage of hires in a unit that was above a suggested cutoff score on selection tests.

Selection in team contexts by its nature involves multilevel hypotheses (i.e., attributes of the individuals in a team affect team level outcomes; Morgeson et al. 2012), and some researchers found their data to be consistent with this hypothesis (see Bell 2007, Stajkovic et al. 2009 for meta-analyses). We do have a large and growing body of literature that examines aggregated relationships at various levels, but studies of cross-level relationships are rare. For example, Jiang et al. (2012) provided a review of various human resource practices, including selection, as they relate to organizational outcomes; the studies in the Jiang et al. review involved analyses done at the organizational level. There are good reasons why results from individual-level and unit- or organization-level analyses do not coincide. One possible explanation is the failure to consider cross-level differences, as discussed by Ployhart (2012, pp. 671–72). Selection researchers can and

should continue their work to understand the nature, implications, and appropriateness of aggregation and inferences across levels of analysis.

WEB-BASED TESTING

Certainly there has been no greater influence on the delivery and use of tests than technology. Online systems exist that will take the application of a job candidate, deliver online tests, provide feedback to the prospective employer and/or the candidate, do background checks, conduct interviews, and track the candidate once he or she enters a firm. Although these systems are obviously efficient and may allow the assessment of constructs that would be difficult to assess in other modes (e.g., spatial visualization), they raise important concerns about the security of tests and whether test takers may have cheated in some way (e.g., having someone else help take the test or having books or other source material available to find answers). Such concerns have led to a variety of security practices and to verification testing, as well as to studies to examine the equivalence of test scores obtained in different ways. Several recent reviews of the potential and liabilities of and research on automated testing are available (Reynolds & Dickter 2010, Scott & Lezotte 2012, Tippins et al. 2006, Wunder et al. 2010).

One significant concern regarding the introduction of Web-based testing is the equivalence of these tests to proctored computer-based testing or paper-and-pencil measures. Mead & Drasgow (1993) reported that computer and paper-and-pencil versions of cognitive ability tests produced similar results except when speed was a factor in test performance. Two investigations (Bartram & Brown 2004, Salgado & Moscoso 2003) found that biodata and personality measures displayed similar internal consistency and intercorrelations when measured by computer and paper. Ployhart et al. (2003) found that Web-based proctored tests (as compared with paper-and-pencil tests) had lower means, greater variances, better internal consistency, and greater covariances. These differences were larger for personality tests than for biodata and situational judgment measures. However, none of these studies included unproctored online versions of tests, which have become increasingly popular.

Unproctored Internet testing, in which an examinee takes a test at home or some other convenient location without supervision, has become increasingly popular in the past decade or so and has now become the only method by which some organizations assess large portions of their applicant pool and the primary method of delivery of exams by test publishers and consulting firms. This mode of test delivery is convenient for both employer and employee and provides very fast results inexpensively. However, it produces significant concerns about cheating, test security, and standardization. The benefits and problems associated with unproctored Internet testing have been explored in a series of articles in *Industrial and Organizational Psychology* introduced by Tippins (2009), and the International Testing Commission (2006) has issued a set of guidelines for unproctored testing practices. Perhaps the most common approach to alleviate some of the concerns about these applications is to use unproctored test scores as a screen and then verify the scores with some version of a proctored exam given to a smaller set of examinees. Weiner & Morrison (2009) described one approach in which an unproctored personality test was used as an initial screen with the subsequent administration of a proctored cognitive test. More sophisticated approaches use a computer-adaptive test as an initial screen and follow that with another verification test, also a computer-adaptive test but administered in a proctored setting. The score on the initial test can be used as a prior ability estimate in beginning the second proctored exam, thus shortening the testing process considerably and allowing for a quick determination of the possibility that a person cheated in some way on the unproctored exam.

Internet test use in unproctored and proctored settings has proceeded largely in the absence of research on the equivalence of test scores, their validity, their reliability, and the extent or nature

of examinee cheating and test security. Such research is needed and will almost certainly appear in our journals in the next several years.

SITUATIONAL FRAMING OF TEST STIMULI

To allow for tests to be used in multiple settings without expensive revisions and renorming and revalidation efforts, most tests are general in content and avoid content that would be unique to a particular work setting. However, Schmit et al. (1995) found that contextualizing items of the NEO-PI conscientiousness scale used to predict students' grade point averages increased the validity from .25 to .41 and .46 in two different conditions. The manipulation of context involved the rather simple addition of the words "at work" to each item of the scale. Subsequent work in field settings (e.g., Pace & Brannick 2010) showed superior validity of a context-specific (versus general) measure of openness in the prediction of creative performance. Most convincing of the importance of context is a meta-analysis by Shaffer & Postlethwaite (2012). They found contextualized measures of all the Big Five constructs were superior in validity to noncontextualized measures. The differences in validities ranged from .08 for conscientiousness to .17 for openness and extraversion. These results have important implications for the use of personality measures in the prediction of job performance; a very simple modification to a generic personality measure provides significant improvement in validity.

THE CONTEXT OF SELECTION

Whether ability or personality is related to measures of job performance can also be related to the context in which selection takes place. At first glance this statement may appear inconsistent with the notion that test validities, particularly those for cognitive ability, generalize across situations. However, we can have generalizable relationships at the construct level and still observe relatively large differences across different situations—even in the presence of relatively large N. Certainly the optimal use of selection tests is limited by the context in which they are used. Legal constraints (Gutman 2012, Sackett et al. 2010) are one major constraint in the United States and to some extent in Europe and other parts of the world. The place and hours employees work can determine the nature of selection instruments used and their validity (Bauer et al. 2012). Also, the cultural or national context is a major determinant of the way in which selection takes place and the measurement tools that are acceptable (Steiner 2012, Caligiuri & Paul 2010). The concept of testing and even the acceptance of the notion that individual differences exist vary across cultures and countries. Finally, the degree to which selection is (or is not) integrated with other human resource functions, such as training, performance management, and reward systems, can impact the validity and utility of selection tests. Finally, Tippins (2012) has provided an excellent discussion of the impact of the ways in which selection processes are implemented, including administration, scoring, and use of test results.

MISCELLANEOUS CONTRIBUTIONS

Several developments that do not fit easily in the outline above merit attention. Lievens & DeSoete (2011) described a number of innovative selection techniques that show promise. They mentioned contextualized measures (see above) and serious games that may help organizations improve their image and, in some instances, the validity of measures. They pointed to the potential for integrity tests, implicit association tests, and conditional reasoning tests as ways to measure values and maladaptive traits. Conditional reasoning tests appear to the test taker to be measuring

inductive reasoning. Problems are provided with alternative solutions, some of which reflect biases related to a targeted construct. James et al. (2004) presented evidence from 10 studies that indicated an average validity of .44 against behavioral criteria of aggression. Berry et al. (2010), however, found a meta-analytic validity between .10 and .16 for the commercially available measure of aggression across a broader set of studies.

Instead of examining self-reports of the Big Five measures, Oh et al. (2011) examined the meta-analytic validity of observer ratings of personality. They found that observer ratings had higher validity (ranging from .18 to .32 corrected) than did self-ratings. Moreover, the observer ratings displayed incremental validity over self-ratings, although the reverse was not true. As pointed out by the authors, these conclusions were based on relatively few studies and participants (about 14 studies and 2,000 participants).

In an interesting and novel examination of the role of personality at work and the influence of the work situation on the expression of personality, Huang & Ryan (2011) found that the expression of conscientiousness was correlated with the immediacy of a task and that extraversion and agreeableness were correlated with the friendliness of customers in a customer service job. The investigators used experience sampling methods to assess personality states at work as well as the work circumstances at that time. Their work supports the notion that personality states at work vary meaningfully within person and that this variation is related to situational factors.

Van Iddekinge and colleagues have reevaluated the role of interest measurement in the prediction of job performance and turnover. For example, in a meta-analysis, Van Iddekinge et al. (2011b) found that interests were related to job performance (.14), training performance (.26), turnover intentions (−.19), and actual turnover (−.15). They also found that validity was higher when the interest measure related to the target job (another form of contextualization) and when a combination of interests was used to predict job performance. In an empirical study, Van Iddekinge et al. (2011a) developed an interest measure targeted to military jobs. Estimated cross-validity of multiple correlations of interest measures with five criteria ranged from .13 to .31. These interest measures also provided incremental validity relative to personality and Armed Forces Qualification Test scores.

SUMMARY

There are a very large number of studies, meta-analyses, and reviews that document the validity of cognitive ability and personality tests and their combination as predictors of work performance. Some authors are critical of the personnel selection field, given the magnitude of the validities reported. However, if one considers the complexity of the job performance phenomena and the organizational constraints on performance and our ability to define and measure performance, the size of the coefficients actually represents one of the most remarkable achievements of psychology. Some of the reasons for these relatively low correlations are enumerated in an article by Cascio & Aguinis (2008). The effect of range restriction and criterion unreliability on the estimation of validity are well documented (Le et al. 2006). Rather than recommending more validation research (or meta-analyses) like the studies that are the focus of the meta-analyses cited in this review, I believe well-planned, well-executed, large-scale international validation studies should be performed (see Schmitt & Fandre 2008 for a description of this type of study).

At this time, however, the literature on cognitive ability and personality provides several solid conclusions for those engaged in the practice of psychology:

1. Cognitive ability measures should predict performance outcomes in most, if not all, jobs and situations.

2. Personality measures should also predict performance, albeit perhaps less well and restricted to those situations in which job analyses or theory support their relevance.

3. Tailor-made, context-specific personality measures, subfacets of the Big Five, and combinations of relevant personality constructs may yield superior validity to the Big Five measures.

4. Rather than rely only on a traditional criterion-related validation study or evidence, practitioners should also use other types of studies and evidence to support test use.

5. When relevant, performance outcomes other than task performance should be used as targets of predictor studies, and care should be taken to match the nature of predictor and criterion constructs.

6. In making decisions about alternative selection procedures, practitioners should consider models of the trade-offs between performance outcomes and the diversity of those hired.

7. Perceptions of test takers and users should be considered in the selection and construction of measures. These perceptions certainly determine the likelihood of the use of the tests and may affect their validity as well.

8. Continued caution regarding the existence of faking on noncognitive measures is warranted. Faking and cheating on knowledge and cognitive ability measures administered in nonproctored Web-based environments do occur. Efforts to minimize these occurrences and remedy their impacts on measurement must continue.

9. Finally, relationships between predictors and outcomes at different levels of analysis should be explored. Aside from potential theoretical implications, observation of such relationships is likely to have an impact on organizational decision makers who are more concerned about organizational or unit outcomes than individual outcomes.

There continues to be research on a host of issues relevant to the evaluation of selection procedures as well as their optimal use in a variety of organizational circumstances. Practitioners and researchers need to be aware of these developments if they want to optimize the utility of their selection systems for organizational performance, workforce diversity, and the reactions of their users and the people to whom these techniques are applied. The research on personnel selection procedures also contributes to our understanding of the structures of ability and personality, as well as of the ways in which these individual differences interact to influence behavior at multiple levels of analysis. It is my hope that, in this review, I have directed the reader to papers that will provide useful answers to questions about human performance in organizations.

DISCLOSURE STATEMENT

LITERATURE CITED

AERA, APA, NCME (Am. Educ. Res. Assoc., Am. Psychol. Assoc., Natl. Counc. Meas. Educ.). 1999. *Standards for Educational and Psychological Testing*. Washington, DC: AERA, APA, NCME

Aguinis H, Culpepper SA, Pierce CA. 2010. Revival of test bias research in preemployment testing. *J. Appl. Psychol.* 95:648–80

Anderson N, Salgado JF, Hulsheger UR. 2010. Applicant reactions in selection: comprehensive meta-analysis into reaction generalization versus situational specificity. *Int. J. Sel. Assess.* 18:291–304

Arthur W Jr, Glaze RM, Villado AJ. 2009. Unproctored Internet-based tests of cognitive ability and personality: magnitude of cheating and response distortion. *Ind. Organ. Psychol.* 2:39–45

Barrick MR, Mount MK. 1991. The Big Five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* 44:1–26

Barrick MR, Mount MK. 1993. Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *J. Appl. Psychol.* 78:111–18

Barrick MR, Mount MK. 2012. Nature and use of personality in selection. See Schmitt 2012, pp. 225–51

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? *Int. J. Sel. Assess.* 9:9–30

Barrick MR, Mount MK, Li N. 2013. The theory of purposeful behavior: the role of personality, higher-order goals, and job characteristics. *Acad. Manag. Rev.* 38:132–53

Barrick MR, Mount MK, Strauss JP. 1993. Conscientiousness and performance of sales representatives. *J. Appl. Psychol.* 78:715–22

Bartram D, Brown A. 2004. Online testing: mode of administration and the stability of OPQ 32i scores. *Int. J. Sel. Assess.* 12:278–84

Bauer TN, Truxillo DM, Mansfield LR, Erdogan B. 2012. Contingent workers: Who are they and how can we select them for success? See Schmitt 2012, pp. 865–78

Bell ST. 2007. Deep-level composition variables as predictors of team performance: a meta-analysis. *J. Appl. Psychol.* 92:595–615

Bergner S, Neubauer AC, Kreuzthaler A. 2010. Broad and narrow personality traits for predicting managerial success. *Eur. J. Work Organ. Psychol.* 19:177–99

Berry CM, Clark MA, McClure TK. 2011. Racial/ethnic differences in the criterion-related validity of cognitive ability tests. *J. Appl. Psychol.* 96:881–906

Berry CM, Ones DS, Sackett PR. 2007. Interpersonal deviance, organizational deviance, and their common correlates: a review and meta-analysis. *J. Appl. Psychol.* 92:410–24

Berry CM, Sackett PR, Tobares V. 2010. A meta-analysis of conditional reasoning tests of aggression. *Pers. Psychol.* 63:361–84

Birkeland SA, Manson TM, Kisamore JL, Brannick MT, Smith MA. 2006. A meta-analytic investigation of job applicant faking on personality measures. *Int. J. Sel. Assess.* 14:317–35

Borman WC, Motowidlo SJ. 1993. Expanding the criterion domain to include elements of contextual performance. In *Personnel Selection in Organizations*, ed. N Schmitt, WC Borman, pp. 71–98. San Francisco: Jossey-Bass

Caligiuri P, Paul KB. 2010. Selection in multinational organizations. See Farr & Tippins 2010, pp. 781–800

Campbell JP, McCloy RA, Oppler SH, Sager CE. 1993. A theory of performance. In *Personnel Selection in Organizations*, ed. N Schmitt, WC Borman, pp. 35–70. San Francisco: Jossey-Bass

Campbell LSK, Catano VM. 2004. Using measures of specific abilities to predict training performance in Canadian forces operating occupations. *Mil. Psychol.* 16:183–201

Carretta TR, Ree MJ. 2000. General and specific cognitive and psychomotor abilities in personnel selection: the prediction of training and job performance. *Int. J. Sel. Assess.* 8:227–36

Carroll JB. 1993. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York: Cambridge Univ. Press

Cascio WF, Aguinis H. 2008. Staffing twenty-first-century organizations. *Acad. Manag. Ann.* 2:133–65

Casillas A, Robbins S, McKinniss T, Postlethwaite B, Oh IS. 2009. Using narrow facets of an integrity test to predict safety: a test validation study. *Int. J. Sel. Assess.* 17:119–25

Chiaburu DS, Oh IS, Berry CM, Li N, Gardner RG. 2011. The five-factor model of personality traits and organizational citizenship behavior. *J. Appl. Psychol.* 96:1140–66

Coward WM, Sackett PR. 1990. Linearity of ability-performance relationships: a reconfirmation. *J. Appl. Psychol.* 75:297–300

Dalal RS. 2005. A meta-analysis of the relationship between organizational citizen behavior and counterproductive work behavior. *J. Appl. Psychol.* 90:1241–55

DeCorte W, Lievens F, Sackett PR. 2007. Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *J. Appl. Psychol.* 92:1380–93

DeCorte W, Sackett PR, Lievens F. 2010. Selecting predictor subsets: considering validity and adverse impact. *Int. J. Sel. Assess.* 18:260–70

DeCorte W, Sackett PR, Lievens F. 2011. Designing pareto-optimal selection systems: formalizing the decisions required for selection system development. *J. Appl. Psychol.* 96:907–26

Ellingson JE, Heggestad ED, Makarius EE. 2012. Personality retesting for managing intentional distortion. *J. Personal. Soc. Psychol.* 102:1063–76

Fan J, Gao D, Carroll SA, Lopez FJ, Tian TS, Meng H. 2012. Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *J. Appl. Psychol.* 97:866–80

Farr JL, Tippins NT. 2010. *Handbook of Employee Selection*. New York: Routledge

Frei RL, McDaniel MA. 1998. Validity of customer service measures in personnel selection: a review of criterion and construct evidence. *Hum. Perform.* 11:1–27

Gilliland SW, Steiner DD. 2012. Applicant reactions to testing and selection. See Schmitt 2012, pp. 629–66

Goffin RD, Christiansen ND. 2003. Correcting personality tests for faking: a review of popular personality tests and initial survey of researchers. *Int. J. Sel. Assess.* 11:340–44

Gutman A. 2012. Legal constraints on personnel selection decisions. See Schmitt 2012, pp. 686–720

Harrison DA, Newman DA. 2012. Absence, lateness, turnover, and retirement: narrow and broad understandings of withdrawal and behavioral engagement. In *Handbook of Psychology*, Vol.12, *Industrial and Organizational Psychology*, ed. N Schmitt, S Highhouse, pp. 262–91. New York: Wiley

Hastings SE, O'Neill TA. 2009. Predicting workplace deviance using broad versus narrow personality variables. *Personal. Individ. Differ.* 47:289–93

Hattrup K. 2012. Using composite predictors in personnel selection. See Schmitt 2012, pp. 297–319

Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty-Gerrard MO. 2007. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. *J. Appl. Psychol.* 92:373–85

Hoffman BJ, Dilchert S. 2012. A review of citizenship and counterproductive behaviors in organizational decision-making. See Schmitt 2012, pp. 543–69

Hogan J, Barrett P, Hogan R. 2007. Personality measurement, faking, and employment selection. *J. Appl. Psychol.* 92:1270–85

Hopwood CJ, Wright AGC, Donnellan MB. 2011. Evaluating the evidence for the general factor of personality across multiple inventories. *J. Res. Personal.* 45:468–78

Hough LM. 1992. The "Big Five" personality variables—construct confusion: description versus prediction. *Hum. Perform.* 5:139–55

Hough LM, Dilchert S. 2010. Personality: its measurement and validity. See Farr & Tippins 2010, pp. 299–320

Hough LM, Ones DS. 2001. The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In *Handbook of Work Psychology*, ed. NR Anderson, DS Ones, HK Sinangil, C Viswesvaran, pp. 233-377. New York: Sage

Hough LM, Oswald FL. 2005. They're right, well...mostly right: research evidence. *Hum. Perform.* 18:373–88

Hough LM, Schneider RJ. 1996. Personality traits, taxonomies, and applications in organizations. In *Individual Differences and Behavior in Organizations*, ed. K Murphy, pp. 31–88. San Francisco: Jossey-Bass

Huang JL, Ryan AM. 2011. Beyond personality traits: a study of personality states and situational contingencies in customer service jobs. *Pers. Psychol.* 64:451–88

Hulsheger UR, Maier GW, Stumpp T. 2007. Validity of general mental ability for the prediction of job performance and training success in Germany: a meta-analysis. *Int. J. Sel. Assess.* 15:3–18

Hunter JE, Schmidt FL, Hunter R. 1979. Differential validity of employment tests by race: a comprehensive review and analysis. *Psychol. Bull.* 86:721–35

Int. Test. Comm. 2006. International guidelines on computer-based and internet delivered testing. *Int. J. Test.* 6:143–72

James LR, McIntyre MD, Glisson CA, Bowler JL, Mitchell TR. 2004. The conditional reasoning measurement system for aggression: an overview. *Hum. Perform.* 17:271–95

Jiang K, Lepak DP, Hu J, Baer JC. 2012. How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. *Acad. Manag. J.* 55:1264–94

Johnson JW, Carter GW. 2010. Validating synthetic validation: comparing traditional and synthetic validity coefficients. *Pers. Psychol.* 63:755–95

Johnson JW, LeBreton JM. 2004. History and use of relative importance indices in organizational research. *Organ. Res. Methods* 7:238–57

Johnson JW, Steel P, Scherbaum CA, Hoffman CC, Jeanneret PR, Foster J. 2010. Validations like motor oil: Synthetic is better. *Ind. Organ. Psychol.* 3:305–28

Judge TA, Bono JE. 2001. Relationship of core self-evaluations traits, self-esteem, generalized self-efficacy, locus of control, and emotional stability with job satisfaction and job performance: a meta-analysis. *J. Appl. Psychol.* 86:80–92

Judge TA, Jackson CL, Shaw JC, Scott BA, Rich BL. 2007. Self-efficacy and work-related performance: the integral role of individual differences. *J. Appl. Psychol.* 2:107–27

Kehoe JF, Murphy KR. 2010. Current concepts of validity, validation, and generalizability. See Farr & Tippins 2010, pp. 99–125

Klein KJ, Kozlowski SWJ, eds. 2000. *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions.* San Francisco: Jossey-Bass

Lang JWB, Kersting M, Hülsheger UR, Lang J. 2010. General mental ability, narrower cognitive abilities, and job performance: the perspective of the nested-factors model of cognitive abilities. *Pers. Psychol.* 63:595–640

Le H, Oh I, Robbins SB, Ilies R, Holland E, Westrick P. 2011. Too much of a good thing: curvilinear relationships between personality traits and job performance. *J. Appl. Psychol.* 96:113–33

Le H, Schmidt FL. 2006. Correcting for indirect range restriction in meta-analysis: testing a new meta-analytic procedure. *Psychol. Methods* 11:416–38

LeBreton JM, Ployhart RE, Ladd RT. 2004. A Monte Carlo comparison of relative importance methodologies. *Organ. Res. Methods* 7:258–82

Lievens F, DeSoete B. 2011. Instruments for personnel selection in the 21st century: research and practice. *Gedrag Organ.* 24:18–42

Marcus B, Goffin RD, Johnson NG, Rothstein MG. 2007. Personality and cognitive ability as predictors of typical and maximum managerial performance. *Hum. Perform.* 20:275–85

Mattern KD, Patterson BF. 2013. Test of slope and intercept bias in college admissions: a response to Aguinis, Culpepper, and Price. *J. Appl. Psychol.* 98:134–47

McPhail M. 2007. *Alternative Validation Strategies.* New York: Wiley

Mead A, Drasgow F. 1993. Effects of administration medium: a meta-analysis. *Psychol. Bull.* 114:449–58

Meyer RD, Dalal RS, Bonaccio S. 2009. A meta-analytic investigation into the moderating effects of situational strength on the conscientiousness-performance relationship. *J. Organ. Behav.* 30:1077–102

Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N. 2007. Reconsidering the use of personality tests in personnel selection contexts. *Pers. Psychol.* 60:683–729

Morgeson FP, Humphrey SE, Reeder MC. 2012. Team selection. See Schmitt 2012, pp. 832–48

Mount MK, Oh IS, Burns M. 2008. Incremental validity of perceptual speed and accuracy over general mental ability. *Pers. Psychol.* 61:113–39

Murphy KR, Dzieweczynski JL. 2005. Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Hum. Perform.* 18:343–57

Murphy KR, Shiarella A. 1997. Implications of the multidimensional nature of job performance for the validity of selection tests: multivariate frameworks for studying test validity. *Pers. Psychol.* 50:823–54

Newman DA, Lyon JS. 2009. Recruitment efforts to reduce adverse impact: targeted recruiting for personality, cognitive ability, and diversity. *J. Appl. Psychol.* 94:298–317

Nisbett RE, Aronson J, Blair C, Dickens W, Flynn J, et al. 2012. Intelligence new findings and theoretical developments. *Am. Psychol.* 67:130–59

Oh IS, Wang G, Mount MK. 2011. Validity of observer ratings of the five-factor model of personality traits: a meta-analysis. *J. Appl. Psychol.* 96:762–73

Ones DS, Dilchert S, Viswesvaran C. 2007. In support of personality assessment in organizational settings. *Pers. Psychol.* 60:995–1028

Ones DS, Dilchert S, Viswesvaran C. 2012. Cognitive abilities. See Schmitt 2012, pp. 179–224

Ones DS, Dilchert S, Viswesvaran C, Salgado JF. 2010. Cognitive abilities. See Farr & Tippins 2010, pp. 255–75

Organ DW. 1997. Organizational citizenship behavior: It's construct clean-up time. *Hum. Perform.* 10:85–97

Pace VL, Brannick MT. 2010. Improving prediction of work performance through frame-of-reference consistency: empirical evidence using openness to experience. *Int. J. Sel. Assess.* 18:230–35

Paunonen SV, Ashton MA. 2001. Big Five factors and facets in the prediction of behavior. *J. Personal. Soc. Psychol.* 81:524–39

Penney LM, David EM, Witt LA. 2011. A review of personality and performance: identifying boundaries, contingencies, and future research directions. *Hum. Resour. Manag. Rev.* 21:297–310

Perry SJ, Hunter EM, Witt LA, Harris KJ. 2010. P = f (conscientiousness × ability): examining the facets of conscientiousness. *Hum. Perform.* 23:343–60

Ployhart RE. 2004. Organizational staffing: a multilevel review, synthesis, and model. *Res. Pers. Hum. Resour. Manag.* 23:121–76

Ployhart RE. 2012. Multilevel selection and the paradox of sustained competitive advantage. See Schmitt 2012, pp. 667–85

Ployhart RE, Holz BC. 2008. The diversity-validity dilemma: strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Pers. Psychol.* 61:153–72

Ployhart RE, Moliterno TP. 2011. Emergence of the human capital resource: a multilevel model. *Acad. Manag. Rev.* 3:127–50

Ployhart RE, Van Iddekinge CH, MacKenzie WI Jr. 2011. Acquiring and developing human capital in service contexts: the interconnectedness of human capital resources. *Acad. Manag. J.* 54:353–68

Ployhart RE, Weekley JA, Holtz BC, Kemp C. 2003. Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata, and situational judgment tests comparable? *Pers. Psychol.* 56:733–52

Postlethwaite B, Robbins S, Rickerson J, McKinniss T. 2009. The moderation of conscientiousness by cognitive ability when predicting workplace safety behavior. *Personal. Individ. Differ.* 47:711–16

Pulakos ED, Mueller-Hanson RA, Nelson JK. 2012. Adaptive performance and trainability as criteria in selection research. See Schmitt 2012, pp. 595–613

Pulakos ED, Schmitt N, Dorsey DW, Hedge JW, Borman WC. 2002. Predicting adaptive performance: further tests of a model of adaptability. *Hum. Perform.* 15:299–323

Ree MJ, Carretta TR. 2002. g2K. *Hum. Perform.* 15:3–24

Reeder MC, Ryan AM. 2011. Methods for correcting faking. In *New Perspectives on Faking in Personality Assessment*, ed. M Ziegler, C McCann, RD Roberts, pp. 131–50. New York: Oxford Univ. Press

Reeves CL, Hakel MD. 2002. Asking the right questions. *Hum. Perform.* 15:47–74

Reynolds DH, Dickter DV. 2010. Technology and employee selection. See Farr & Tippins 2010, pp. 171–94

Roth PL, Switzer FS III, Van Iddekinge CH, Oh I-S. 2011. Toward better meta-analytic matrices: how input values can affect research conclusions in human resource management simulations. *Pers. Psychol.* 64:899–936

Rotundo M, Spector PE. 2010. Counterproductive work behavior and withdrawal. See Farr & Tippins 2010, pp. 489–512

Ryan AM, Sacco JM, McFarland LA, Kriska SD. 2000. Applicant self-selection: correlates of withdrawal from a multiple hurdle process. *J. Appl. Psychol.* 85:163–79

Rynes SL, Colbert AE, Brown KG. 2002. HR professionals' beliefs about effective human resource practices: correspondence between research and practice. *Hum. Resour. Manag.* 41:149–79

Sackett PR, DeCorte W, Lievens F. 2008. Pareto-optimal predictor composite formation: a complementary approach to alleviating the selection quality/adverse impact dilemma. *Int. J. Sel. Assess.* 16:206–9

Sackett PR, Gruys ML, Ellingson JE. 1998. Ability-personality interactions when predicting job performance. *J. Appl. Psychol.* 3:545–56

Sackett PR, Laczo RM, Lippe ZP. 2003. Differential prediction and the use of multiple predictors: the omitted variables problem. *J. Appl. Psychol.* 88:1046–56

Sackett PR, Lievens F. 2008. Personnel selection. *Annu. Rev. Psychol.* 59:419–50

Sackett PR, Putka DJ, McCloy RA. 2012. The concept of validity and the process of validation. See Schmitt 2012, pp. 91–118

Sackett PR, Shen W, Myors B, Lievens F, Schollaert E, et al. 2010. Perspectives from twenty-two countries on the legal environment. See Farr & Tippins 2010, pp. 651–76

Sackett PR, Yang J. 2000. Correction for range restriction: an expanded typology. *J. Appl. Psychol.* 85:112–18

Salgado JF, Anderson N, Moscoso S, Bertua C, de Fruyt F. 2003a. International validity generalization of GMA and cognitive abilities: a European community meta-analysis. *Pers. Psychol.* 56:571–605

Salgado JF, Anderson N, Moscoso S, Bertua C, de Fruyt F, Rolland JP. 2003b. A meta-analytic study of general mental ability validity for different occupations in the European community. *J. Appl. Psychol.* 88:1068–81

Salgado JF, Moscoso S. 2003. Internet-based personality testing: equivalence of measures and assessees' perceptions and reactions. *Int. J. Sel. Assess.* 11:194–203

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. *Psychol. Bull.* 124:262–74

Schmidt FL, Oh I, Le H. 2006. Increasing the accuracy of corrections for range restriction: implications for selection procedure validities and other research results. *Pers. Psychol.* 59:281–305

Schmit MJ, Ryan AM. 1997. Applicant withdrawal: the role of test-taking attitudes and racial differences. *Pers. Psychol.* 50:855–76

Schmit MJ, Ryan AM, Stierwalt SL, Powell AB. 1995. Frame-of-reference effects on personality scale scores and criterion-related validity. *J. Appl. Psychol.* 80:607–20

Schmitt N, ed. 2012. *The Oxford Handbook of Personnel Selection and Assessment*. New York: Oxford Univ. Press

Schmitt N, Fandre J. 2008. The validity of current selection methods. In *Oxford Handbook of Personnel Psychology*, ed. S Cartwright, CL Cooper, pp. 163–93. New York: Oxford Univ. Press

Schmitt N, Oswald FL. 2006. The impact of corrections for faking on the validity of noncognitive measures in selection settings. *J. Appl. Psychol.* 91:613–21

Scott JC, Lezotte DV. 2012. Web-based assessments. See Schmitt 2012, pp. 485–513

Shaffer JA, Postlethwaite BE. 2012. A matter of context: a meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Pers. Psychol.* 65:445–94

SIOP (Soc. Ind. Organ. Psychol.). 2003. *Principles for the Validation and Use of Personnel Selection Procedures*. Bowling Green, OH: SIOP

Stajkovic AD, Lee D, Nyberg AJ. 2009. Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of a mediation model. *J. Appl. Psychol.* 94:814–28

Stark S, Chernyshenko OS, Drasgow F, White LA. 2012. Adaptive testing with multidimensional pairwise preference items: improving the efficiency of personality and other noncognitive assessments. *Organ. Res. Methods* 15:463–88

Steiner DD. 2012. Personnel selection across the globe. See Schmitt 2012, pp. 740–67

Tam AP, Murphy KR, Lyall JT. 2004. Can changes in differential dropout rates reduce impact? A computer simulation study of a multi-wave selection system. *Pers. Psychol.* 57:905–34

Tett RP, Christiansen ND. 2007. Personality tests at the crossroads: a response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Pers. Psychol.* 60:967–94

Thoresen CJ, Bradley JC, Bliese P, Thoresen JD. 2004. The Big Five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *J. Appl. Psychol.* 89:835–53

Thurstone LL. 1938. *Primary Mental Abilities*. Chicago: Univ. Chicago Press

Tippins NT. 2009. Internet alternatives to traditional proctored testing: Where are we now? *Ind. Organ. Psychol.* 2:2–10

Tippins NT. 2012. Implementation issues in employee selection testing. See Schmitt 2012, pp. 881–902

Tippins NT, Beaty J, Drasgow F, Gibson WM, Pearlman K, et al. 2006. Unproctored internet testing in employment settings. *Pers. Psychol.* 59:189–225

Truxillo DM, Bodner TE, Bertolino M, Bauer TN, Yonce CA. 2009. Effects of explanation on applicant reactions: a meta-analytic review. *Int. J. Sel. Assess.* 17:346–61

Van der Linden D, te Nijenhuis JT, Bakker AB. 2010. The general factor of personality: a meta-analysis of Big Five intercorrelations and a criterion-related validity study. *J. Res. Personal.* 44:315–27

Van Hooft EAJ, Born MP. 2012. Intentional response distortion on personality tests: using eye-tracking to understand responses processes when faking. *J. Appl. Psychol.* 97:301–16

Van Iddekinge CH, Ferris GR, Perrewe PL, Perryman AZ, Blass FR, Heetderks TD. 2009. Effects of selection and training on unit-level performance over time: a latent growth modeling approach. *J. Appl. Psychol.* 94:829–43

Van Iddekinge CH, Ployhart RE. 2008. Developments in the criterion-related validation of selection procedures: a critical review and recommendations for practice. *Pers. Psychol.* 61:871–925

Van Iddekinge CH, Putka DJ, Campbell JP. 2011a. Reconsidering vocational interests for personnel selection: the validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *J. Appl. Psychol.* 96:13–33

Van Iddekinge CH, Roth PL, Putka DJ, Lanivich SE. 2011b. Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *J. Appl. Psychol.* 96:1167–94

Van Iddekinge CH, Roth PL, Raymark PH, Odle-Dusseau HN. 2012. The criterion-related validity of integrity tests: an updated meta-analysis. *J. Appl. Psychol.* 97:499–530

Van Iddekinge CH, Taylor MA, Eidson CE Jr. 2005. Broad versus narrow facets of integrity: predictive validity and subgroup differences. *Hum. Perform.* 18:151–77

Weiner JA, Morrison JD. 2009. Unproctored online testing: environmental conditions and validity. *Ind. Organ. Psychol.* 2:27–30

Whetzel DL, McDaniel MA, Yost AP, Kim N. 2010. Linearity of personality-performance relationships: a large-scale examination. *Int. J. Sel. Assess.* 18:310–20

Witt LA. 2002. The interactive effects of extraversion and conscientiousness on performance. *J. Manag.* 28:835–51

Wunder RS, Thomas LL, Luo Z. 2010. Administering assessments and decision making. See Farr & Tippins 2010, pp. 377–98

Ziegler M. 2011. Applicant faking: a look into the black box. *Ind.-Organ. Psychol.* 49:29–37

Ziegler M, McCann C, Roberts RD, eds. 2011. *New Perspectives on Faking in Personality Assessment*. New York: Oxford Univ. Press

Zyphur MJ, Bradley JC, Landis RS, Thoresen CJ. 2008. The effects of cognitive ability and conscientiousness on performance over time: a latent growth model. *Hum. Perform.* 21:1–27

**Annual Review of
Organizational
Psychology and
Organizational Behavior**

Volume 1, 2014

Contents

Errata

An online log of corrections to *Annual Review of Organizational Psychology and
Organizational Behavior* articles may be found at http://www.annualreviews.org/
errata/orgpsych.

ANNUAL REVIEWS

It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Statistics and Its Application

Volume 1 • Online January 2014 • http://statistics.annualreviews.org

Editor: **Stephen E. Fienberg,** *Carnegie Mellon University*

Associate Editors: **Nancy Reid,** *University of Toronto*

Stephen M. Stigler, *University of Chicago*

The *Annual Review of Statistics and Its Application* aims to inform statisticians and quantitative methodologists, as well as all scientists and users of statistics about major methodological advances and the computational tools that allow for their implementation. It will include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

Complimentary online access to the first volume will be available until January 2015.

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.

ANNUAL REVIEWS | Connect With Our Experts

Tel: 800.523.8635 (US/CAN) | Tel: 650.493.4400 | Fax: 650.424.0910 | Email: service@annualreviews.org