HUMAN PERFORMANCE, *15*(1/2), 211–231

Agreements and Disagreements on the Role of General Mental Ability (GMA) in Industrial, Work, and Organizational Psychology

Chockalingam Viswesvaran
Department of Psychology
Florida International University

Deniz S. Ones
Department of Psychology
University of Minnesota

General Mental Ability (GMA) has empirical evidence supporting it as a strong predictor of job performance. However, there are agreements and disagreements about the role of GMA in Industrial, Work, and Organizational (IWO) psychology. Some embrace it enthusiastically; some tolerate it; some spend their entire careers looking for ways to minimize the effects of GMA in personnel selection; and, finally, some revile and loath the very concept. The reasons for the divergence vary, and in this special issue of *Human Performance* we brought together leading IWO psychologists and researchers to discuss the potential role of GMA in personnel selection. In this summary, we synthesize, around eight themes, the main points of agreement and disagreements across the contributing authors. The major themes and questions are: (a) predictive value of GMA for real-life outcomes and work behaviors, (b) predictive value of GMA versus specific abilities, (c) the consequences of the criterion problem for GMA validities, (d) is utility evidence for GMA convincing?, (e) are the negative reactions to GMA tests a result of group differences?, (f) is theoretical knowledge of GMA adequate?, (g) is there promise in new methods of testing for GMA?, and (h) what is the current status of non-GMA predictors as substitutes or supplements to GMA?

Requests for reprints should be sent to Chockalingam Viswesvaran, Department of Psychology, Florida International University, Miami, FL 33199. E-mail vish@fiu.edu or Deniz S. Ones, Department of Psychology, University of Minnesota, 75 East River Rd., Minneapolis, MN 55455–0344. E-mail deniz.s.ones-1@tc.umn.edu

Intelligence has been a much debated construct in all of its history. Some swear by it, others swear at it. This is true not only in the area of personnel selection, but across other fields where it has been employed (Jensen, 1980).

Cognitive ability has been used as a predictor in personnel selection for well over 80 years. Although a voluminous literature exists on the topic of cognitive ability, suggesting that it is at least an extensively (if not thoroughly) researched topic, a reading of this (sometimes) contentious literature leads to the conclusion that it may not be well-understood and that there exists less than desirable agreement among Industrial, Work, and Organizational (IWO) psychologists about the appropriate role of cognitive ability in IWO psychology in general, and personnel selection in particular.

Of interest, some of the issues that are debated in this issue about the use and abuse of General Mental Ability (GMA) tests have been around even before intelligence tests were created. The articles in this special issue mirror the literature at large. In any topic of interest to scientists (and especially IWO psychologists), disagreements exist. Consensus omminum is antithetical to science; if everyone agrees on a postulate, no one is likely to conduct research on that postulate. Thus, although some disagreement is expected in any topic, for a topic such as cognitive ability that has played a central role in the practice of IWO psychology for so many years, we would have expected broader agreement.

In this special issue, we brought together distinguished scientists who have grappled with some of the most fundamental issues surrounding the utilization of *g* in IWO psychology for many years. Of interest, they also disagree with each other on a number of fundamental points. Agreements also exist, but they are few. Even the labels used for the construct are not constant. In the set of 11 articles included in this issue, we see terms such as *g,* psychometric-*g,* general mental ability, cognitive ability tests, general cognitive ability, and of course, intelligence tests. In our summary, we refer to this construct as cognitive ability or GMA.

The agreements and disagreements surrounding the use of cognitive ability in selection have to do with both the conclusions that can be drawn from the empirical research conducted on the construct and its measures as well as judgments about the moral obligations IWO psychologists have to society at large. For example, some (see Goldstein, Zedeck, & Goldstein, this issue) question the reliance on GMA not only because of the group differences that result, but also on grounds that it is not a good (enough) predictor. Others level criticisms against reliance on GMA as a predictor because its use results in adverse impact against minorities (Kehoe, this issue).

In this summary, we organize our discussion around eight themes and questions. We underscore the differing opinions expressed on these eight issues by our contributors: their differences on how we should proceed. We hope this summary will enable readers to have an overview of the issues debated as well as the different stances on each issue. The eight themes and questions occurring across

the different articles are as follows: (a) predictive value of GMA for real-life outcomes and work behaviors, (b) predictive value of GMA versus specific abilities, (c) the consequences of the criterion problem for GMA validities, (d) is utility evidence for GMA convincing?, (e) are the negative reactions to GMA tests a result of group differences?, (f) is theoretical knowledge of GMA adequate?, (g) is there promise in new methods of testing for GMA?, and (h) what is the current status of non-GMA predictors as substitutes or supplements to GMA?

PREDICTIVE VALUE OF GMA FOR REAL-LIFE OUTCOMES AND WORK BEHAVIORS

In empirical investigations of cognitive ability, by far the most attention has been devoted to predictive validities of GMA for important work and nonwork outcomes.

Ree and Carretta (this issue) as well as Gottfredson (this issue) summarizes the relations between many life outcomes and GMA scores. Drawing on comprehensive and authoritative summaries (Brand, 1987; Herrnstein & Murray, 1994; Jensen, 1980, 1998; Lubinsky & Humphreys, 1997), Schmidt (this issue) states that GMA is positively related to several life outcomes such as educational level, adult income, and positive health-related behaviors; and negatively related to disciplinary problems, delinquency, and crime rates. Gottfredson (1997) presents substantial correlations (.30–.70, depending on which corrections are made) between GMA scores and everyday activities, such as finding which route in a map.

There is general agreement that there has been more than adequate examination of criterion-related validities of cognitive ability tests for different jobs and in different settings. Schmidt (this issue) summarizes most of the validity generalization studies that have been conducted on GMA in his tables. In the special issue, we even have the first validity generalization evidence for two European countries (Salgado & Anderson, this issue).

Reeve and Hakel (this issue) use an analogy from psychological testing literature to underscore the role of GMA in real-life environments. The probability of getting an item correct on a test is hypothesized to be a monotonically increasing function of an underlying trait. Individuals at a given trait level can answer an item of particular item difficulty with a probability of .50, for that item difficulty (and all items of lesser difficulty with greater probability). It is possible, however, that an individual may fail to answer an easier item. Such discrepancies may occur due to carelessness or may be due to good luck or guessing. Similarly, the cognitive complexity of environments vary. Individuals can tackle environments of complexity at or below their cognitive level. Just as a test-taker may miss an item of low difficulty (compared to his or her trait level), it is possible that an intelligent person may fail to succeed in an environment with lower complexity than his or her individual capacity. Thus, cognitive ability is a necessary but not a sufficient condition for success in the real-life out-

comes. This idea of individuals needing some level of cognitive complexity to survive and thrive in an environment (with a given level of complexity) is also seen in Gottfredson's article in this issue. Future research can address how this (probabilistic) relation between GMA and real-life outcomes is moderated, mediated, or suppressed by other situational variables and individual traits.

Disagreements arise from interpreting the significance of these findings. On one hand, Reeve and Hakel (this issue) state that if there is only one thing we can know of an individual, then knowing the GMA score will result in the best prediction. Kehoe (this issue), Tenopyr (this issue), as well as Murphy (this issue), accepts that GMA has predictive validity, although they differ in their enthusiasm in endorsing its use.

In contrast, other articles question the relative importance of GMA for real-life outcomes. Criticisms revolve around three main points. First, a central criticism, seen in the articles by Goldstein et al. (this issue), Outtz (this issue), as well as Sternberg and Hedlund (this issue), is that the validities of GMA for predicting real-life outcomes are low. Goldstein et al. state that the percentage variance explained is only 25%, which suggests that 75% of the variance in outcomes are not accounted for.[1] A similar concern is echoed by Sternberg and Hedlund. Goldstein et al. go so far as to state that, given the small amount of variance in work performance explained by cognitive ability, the finding of a generalizable relation between GMA and job performance has inhibited progress in IWO psychology. Murphy also states that we might have become victims of our own success (with GMA).

How high should a correlation be before we can consider it to be substantial? Would researchers and practitioners have been happier if GMA predicted job performance with a validity of 1.00? We think not. Such a scenario would mean that the only determinant of performance would be GMA, and other factors such as motivation, hard work, honesty, and integrity would not matter. It would mean that no other variable could compensate for low cognitive ability test scores. It would mean the complete collapse of IWO psychology as a field: No human resources intervention (e.g., training, development, compensation, and other management interventions) could affect relative rankings on job performance, as the sole determinant of criterion would be GMA. This is not a world in which many of us would like to live.

A second disagreement revolves around whether corrected or uncorrected validities should be interpreted as reflective of the true value of cognitive ability tests. The point of disagreement arises from which corrections, if any, should be applied to observed correlations. For example, Outtz (this issue) noted that observed GMA validities reported in the literature are around .30. His argument is that we should focus on

[1]A point worth noting here is that, in gauging the strength of the relation between GMA and performance, the appropriate index of effect size to use is the correlation coefficient (i.e., criterion-related validity) rather than the coefficient of determination, because "evaluating effect size in terms of variance accounted for may lead to interpretations that grossly underestimate the magnitude of a relation" (Ozer, 1985, p. 307).

observed correlations, as these values reflect the state of affairs in practice. The irony of selection is that the more selective one is on the predictor variable, the smaller a correlation one observes with the criterion of interest. If a selection tool has been truly successful, we should observe no or little variance among those selected on that predictor variable. On the other hand, the validity for job applicants (an unselected group) is the value of interest. Range restriction corrections help to gauge the value of the selection tool for applicants (Kuncel, Campbell, & Ones, 1998). Also, if a criterion is unreliably measured, this should not affect judgments about the value of a predictor. Hence, there is a need to correct for unreliability in the criterion as well (Viswesvaran, Ones, & Schmidt, 1996). In passing, we should note that, after deciding to make criterion unreliability and range restriction corrections, another question that arises is which estimates should be used in making them. For example, we recently had an informative exchange on the appropriate reliability coefficient to use for correcting observed validity coefficients for the criterion of supervisory ratings of job performance (Murphy & DeShon, 2000; Schmidt, Viswesvaran, & Ones, 2000). Whether interrater reliabilities, test–retest, or internal consistency reliabilities are used in the corrections will depend on how the criterion construct is conceptualized and measured, as well as on what types of inferences we wish to make.

A third point of disagreement (cf. Sternberg & Hedlund, this issue) is that many psychologists and researchers believe that intelligent behavior is the result of more than just what intelligence tests measure. This may be an educational or public relations issue. Even proponents of GMA validity do not question this view; they merely state that GMA is a major predictor. It is likely that researchers and laypersons use this criticism as a psychological defense mechanism. Given that positive self-delusion is related to well-being, society may need to maintain the viability of this criticism. Future research should investigate how we can promote the use of valid tests without injuring people's self-esteem. Nonetheless, we cannot base research conclusions on what laypersons and researchers believe. That is not science. Many laypersons (and even psychologists?) and researchers in other sciences wonder whether I–O psychology stands for input–output psychology. This misunderstanding does not preclude us from studying industrial–organizational psychology. Thus, misconceptions of laypersons may need to be addressed in educational initiatives and not accepted as truth.

PREDICTIVE VALUE OF GMA VERSUS SPECIFIC ABILITIES

When Ree and colleagues (e.g., Ree & Earles, 1991) titled a series of articles "Not much more than g," a misconception arose about what was being inferred. Many erroneously interpreted "much more than g" as referring to noncognitive mea-

sures. The first thing to note is that specific abilities as referred to here do not include personality and motivational constructs or other popular predictors such as interviews that may be designed to assess these. Thus, we can state that all authors in this special issue agree that there are noncognitive predictors or alternatives that could be used to enhance job performance predictions.

Disagreements arise when we focus on specific abilities such as verbal comprehension, block design, number sequencing, finger dexterity, mechanical knowledge, and so forth. Although any specific definition of specific abilities can be challenged, a referral to Carroll (1993) will make clear what we mean by specific abilities here (the boundaries may be fuzzy, but a perusal of the listed specific abilities in Carroll, 1993, will provide a fairly good idea of what we mean by specific abilities in this context). At issue is whether specific abilities increment the validity of GMA.

The attractiveness of specific abilities lies in the desirability of potential outcomes. If specific abilities are important, then we can have a situation where different individuals will be in the top 10% of different specific abilities. Thus, we can conclude that all individuals are in the top 10%, albeit on different specific abilities. This Lake Wobegone effect is comforting to everyone's ego. Schmidt (this issue) elaborates on this psychological process in his article. Thus, multiple aptitude theory is more palatable to all. Reeve and Hakel (this issue) argued that domain-specific performance requires focused, intensive, and deliberate practice. Therefore, GMA may be a necessary and distal determinant of performance, but specific skills obtained by practice are more proximal determinants.

In contrast to what we may desire or believe about domain-specific performance, empirical evidence suggests that the incremental variance when specific abilities are added to GMA is low. All authors who addressed this issue in their articles here (Schmidt, Kehoe, Tenopyr, and Murphy, as well as Ree and Carretta) agree that there is not much more validity to be gained from specific abilities than *g* (as indexed by incremental validity in the extant literature). However, the authors differ in their evaluation of this empirical finding.

First, Murphy (this issue) states that, although the incremental validity of specific abilities over GMA may be low, it is possible that different people may be selected depending on the specific abilities emphasized. Schmidt (this issue), on the other hand, argues that the military and several police organizations have tried this suggestion, but differential prediction by specific abilities has not been empirically supported. More research on civilian settings (and including a variety of jobs, perhaps across organizations, across industries, and across occupations) will be informative.

Second, Tenopyr (this issue) argues that the low incremental validity of specific abilities may be due to the fact that supervisors have little opportunity to observe specific abilities. The question then arises whether the low opportunity to observe is an accurate reflection of the relative importance of specific abilities or a reflection of criterion deficiency. If the former, then we can dismiss specific abilities as not being important in the workplace; if the latter, we may need to explore methods

to refine our criteria. However, Murphy and Cleveland (1995) reported that raters were more accurate in their global evaluations than in recalling specific behavioral acts. If this finding from performance appraisal literature is any guide, tinkering with the criterion may not easily provide incremental validity for specific abilities.

Third, Kehoe (this issue) argues that the low incremental validity may be due to its computation in a restricted population. He noted that GMA is often validated by using samples from the general applicant population, whereas specific ability tests are administered only in restricted samples. However, this explanation appears to be viable only for some exotic specific abilities; most commonly referred to specific abilities can be administered to everyone in the general population. Furthermore, we can correct for range restriction in a third variable to assess the validity of specific abilities in the general population.

Kehoe (this issue) also raises a point that, although the incremental validity may be low for specific abilities, the use of specific abilities may reduce group differences and result in more positive applicant reactions (presumably because these tests are more face valid). Furthermore, Kehoe notes that it might be easier to legally defend a specific ability test. In light of these two advantages of specific abilities over GMA, Kehoe suggests that specific ability tests should be preferred over tests of GMA. He argues that we should require test publishers to demonstrate the incremental validity of GMA over specific abilities. Ree and Carretta (this issue) countered that one cannot measure specific abilities without measuring GMA. Furthermore, a composite of specific abilities is by definition a measure of GMA.

Unlike specific abilities, *g*-loaded tests are not domain dependent. Measuring GMA enhances generalizability across jobs and criterion facets also. We suspect the general factor observed in job performance measures (see Viswesvaran, 1993; Viswesvaran & Ones, 2000) is partially caused by the GMA that cuts across job performance dimensions as a determinant.

Nevertheless, perhaps another point to note is that empirical findings that relate to the lack of incremental validity of specific abilities over GMA have been limited to a handful of criteria (e.g., overall job performance, training success). We cannot help but wonder if future research will find the same pattern of results for other criteria, such as counterproductive behaviors. The bandwidth and fidelity of what is being predicted is crucial (Ones & Viswesvaran, 1996).

THE CONSEQUENCES OF THE CRITERION PROBLEM FOR GMA VALIDITIES

The criterion problem has been with us for almost a century (Austin & Villanova, 1992; Landy & Farr, 1980). Most research on GMA has demonstrated the predictive validity for training performance, overall job performance, and technical performance. Disagreements exist on the adequacy of these criteria.

Questions are raised about the fact that some of the validation studies have used training performance as the criterion (cf. Goldstein et al., this issue). Because training performance involves the same or similar cognitive tasks as those used in GMA tests, these validity coefficients are called into question. However, Ree and Carretta (this issue) notes that, before anyone can perform the job tasks, they have to learn what to perform and how to perform (declarative and procedural knowledge), rendering training performance a crucially important criterion. Goldstein et al. (this issue) also stresses the distinction between task and contextual performance (Borman & Motowidlo, 1993, 1997) and that GMA, although valid for task performance, may not be as valid as personality variables for predicting contextual performance. Of course, this is countered by Gottfredson (this issue) that no organization would forego selecting individuals for core task performance.

Outtz (this issue) points out that race differences in GMA are greater than race differences in job performance. This is a result of GMA tests having less than perfect predictive validity. Schmidt (this issue) notes that, given the criterion—related validity of GMA for predicting job performance is around .50; a difference of 1 *SD* in GMA should only translate to a difference of .50 *SD* in job performance. If this were not the case, it would actually be an indication of differential prediction and test bias (Rotundo & Sackett, 1999). However, Outtz stresses that even the reduced (0.5 vs. 1.0 *SD*) differences reported in the literature (Ford, Kraiger, & Schechtman, 1986; Kraiger & Ford, 1985) may be misleading, and that actual race differences in job performance may be much less than .50 *SD* units (Pulakos, Oppler, White, & Borman, 1989; Sackett & DuBois, 1991). Unreliability in performance ratings needs to be taken into consideration. As Schmidt points out, there is voluminous research that triangulates the findings from supervisory ratings with those obtained from objective records and work sample tests (e.g., Bommer, Johnson, Rich, Podsakoff, & MacKenzie, 1995; Borman, 1978, 1979; Heneman, 1986; Schmidt & Hunter, 1998).

Goldstein et al. (this issue) also argues that IWO psychologists should not restrict themselves to individual job performance, but should consider team and organizational performance. That is, in addition to showing validity of GMA for predicting individual job performance, we may need to explore the validity of GMA for organizational effectiveness. Of course, utility analysis does just that, but Goldstein et al. (along with Tenopyr, this issue) do not have much confidence in those utility estimates.

IS UTILITY EVIDENCE FOR GMA CONVINCING?

Schmidt (this issue) and Hunter and Schmidt (1996) estimate the utility of GMA as substantial to both organizations and to the economy at large. As Schmidt also notes, the social consequences of ignoring the role of GMA in job performance

may be large (as in the case of police performance). Salgado and Andersen (this issue) provides empirical data suggesting the utility of GMA in European contexts.

However, other articles in this special issue also raise several caveats about utility estimates. For example, Outtz (this issue) wonders how observed correlations of .30 can translate into the gargantuan estimates of utility being paraded. Tenopyr (this issue) notes how even the conservative utility estimates (Schmidt & Hunter, 1981) strain credulity so much that she avoids using such estimates.

It may be worthwhile to remember that the decision theoretic models used to express the utility for tests of GMA are not applicable only to these tests. There is nothing methodologically wrong with the dollar value estimates. Furthermore, even if dollar value estimates are dismissed, other indicators of value (e.g., reduction in workforce to maintain same productivity; training costs; training time reductions; percent increase in output) can provide compelling utility statements. However, utility analyses, as currently carried out, do reductionistically focus on productivity as the primary goal.

Murphy (this issue) maintaines that the statement that we should develop and recommend to employers' selection systems that maximize productivity is a value statement in itself. For example, one can argue that the goal should be to develop a selection system that maximizes opportunities for all groups in society. Outtz (this issue) cites a study by Silva and Jacobs (1993) that suggests that the utility loss from increased minority hiring is not substantial. Outtz suggests that organizations may find a competitive advantage in diversity, especially in a world of increasing diversity.

Future research should explore the stakeholder model that Murphy (this issue) describes in his article. Specifically, research from economics, judgment, and decision making can be used to understand how individuals make utility estimates. The question of which interests to include in a stakeholder model is still an unanswered one. Other issues to consider will have to include whether stakeholders have accurate insights into their own values. Much of the policy capturing studies in human judgment and decision making are not optimistic regarding this question.

ARE THE NEGATIVE REACTIONS TO GMA TESTS A RESULT OF GROUP DIFFERENCES?

In the United States, there are inescapable group differences on cognitive ability measures. In making desirability judgments about using various predictors as part of selection systems, IWO psychologists compare group differences across different predictors. We saw it in this special issue, when various authors argued for using predictors with smaller group differences in lieu of GMA tests. Yet, no mention is made of the fact that if GMA tests are more reliable than other predictors, the advantage of lower group differences predictors might be illusory. Outtz (this issue) states that GMA produces racial differences that are 3 to 5

times that produced by interviews, biodata, and work sample tests. We wonder about the magnitude of group differences masked by measurement error in predictors such as interviews. In a similar vein, Kehoe (this issue) states that specific abilities show smaller group differences than GMA. This may be a case of unreliability and range restriction masking the true nature of group differences on measures of specific abilities (specific abilities may typically be measured in more select samples than GMA). It may be important to know both the observed and the true (corrected) magnitudes of group differences on predictor measures. The former gives an indication of the differences that are likely to be observed in using the test, whereas the latter can provide insights into whether the construct being measured or statistical artifacts (e.g., measurement error and range restriction) are at the root of differences observed.

On another front, in the interest of making progress in our science, future research should be careful about the terminology used. Clarity demands it. For example, it is dramatic to use the word *discrimination*. But what does it mean? As another example, note how Murphy (this issue) argues about the trade-off between efficiency and equity. On reflection, it is clear that he is talking about group and not individual equity, especially if it is defined from an unqualified individualism perspective (Hunter & Schmidt, 1977). Terms such as *group differences, discrimination, adverse impact, bias*, and *fairness*, although related, are distinct. Carefully crafted and used definitions are crucial to science.

Yet another issue is to consider the logic of the four-fifths rule. Mean differences on predictors may or may not translate themselves into adverse impact. Selectivity also matters. As organizations become more selective, the adverse impact resulting from observed group differences increases. The four-fifths rule infers adverse impact when the selection rate for the low scoring group is less than four-fifths the selection rate for the high scoring group. However, this is only an administrative convenience to balance several social claims (Sackett & Ellingson, 1997). Gottfredson (this issue) notes how job complexity varies across jobs, and as a result the extent of adverse impact varies across jobs when GMA is used for selection. Reeve and Hakel (this issue) compares the cognitive complexity of environments to test item difficulty. Given this gradation in complexity across jobs, it might be intriguing to explore the possibility of a sliding adverse impact rule (e.g., infer adverse impact when the ratio of selection ratios is .50 for high complexity jobs, .80 when considering low complexity jobs). Alternatively, we can eschew the tendency to think in dichotomies (whether or not there is any adverse impact). If we decide to avoid dichotomies, then the question raises as to what the plaintiff has to show to demonstrate adverse impact.

Would the controversy surrounding GMA be less intense if there were no group differences? Although it is only academic to consider such hypothetical scenarios, such considerations can help us better understand the nature and the sources of debate. Some of the authors (Murphy, Kehoe, Outtz) view group dif-

ferences as a major drawback of GMA assessment. For example, Murphy (this issue) states that GMA is the best predictor of job performance, but also the predictor with most adverse impact. Kehoe (this issue) asserts that the dilemma stems primarily from three key findings about GMA and work performance: (a) GMA tests are at least as predictive of work performance as other more job-specific tests in the cognitive domain and other assessments outside the cognitive domain, such as personality (Schmidt & Hunter, 1998); (b) specific abilities have little incremental validity; and (c) there are substantial group differences on GMA tests that are not due to bias in the measurement (Wigdor & Garner, 1982). According to Kehoe, the first two conclusions suggest efficient and effective selection strategies in which GMA is a primary component, but the third suggests undesirable social consequences. Similarly, Outtz (this issue) states that GMA predicts job performance, but the controversy comes from the fact that cognitive ability tests produce racial differences that are 3 to 5 times larger than other predictors—such as biodata, personality inventories, and the structured interview—that are valid predictors of job performance. According to Gottfredson (this issue), most of the criticism against GMA is at least partly driven by a motivation to meet the socially desirable outcome of reducing group differences.

In contrast, Schmidt (this issue) argues that the controversy about GMA would have been just as intense even if there were no group differences on GMA measures. Schmidt notes that even students who are not aware of group differences have negative reactions. In relatively homogenous societies, there are also negative reactions to GMA tests (Salgado & Anderson, this issue). Salgado and Anderson (this issue) reports on applicant reactions from European countries, where group differences (especially racial group differences) are not as much a salient issue as they are in the United States. Their review suggests that negative reactions to GMA tests exist in the European countries. Why are there negative feelings toward a high validity predictor?

Reeve and Hakel (this issue) notes how the past abuses of testing for GMA still haunt us. The relation between GMA assessment and social movements such as eugenics has stigmatized the societal view of GMA. Equality for all has been touted as the cornerstone of Western democracy. The curious thing is that GMA can be used to create a society of equal opportunity for all.

Research on applicant reactions in general, and reactions to GMA assessment, can be valuable in discovering the roots of dislike for GMA tests (Rynes, 1993). However, the study of applicant reactions is still in its infancy (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). The existing work focuses on general psychological principles such as perceptions of justice (Gilliland, 1993). Most of the existing research in organizational research has taken the organization's perspective, and very little work has focused on the applicant's perspective (Gilliland, 1995; Schuler, Farr, & Smith, 1993). Most of the existing research focuses on applicants reactions to various selection system characteristics and very

little addresses the question why such reactions occur or what could be done specifically to address or mitigate such concerns.

Goldstein et al. (this issue) and Sternberg and Hedlund (this issue) provids another potential reason for the limited support GMA assessments enjoy among the general public. These authors points out that the laypeople may be convinced that cognitive ability is not the only determinant, or even the most important determinant of intelligent behavior. This line of reasoning may also partially explain Tenopyr's (this issue) point that, although research suggests validity of GMA increases with increasing job complexity, organizations are less likely to emphasize GMA for high-level jobs than they do for lower level jobs. As discussed earlier, face validity probably also comes into play here.

IS THEORETICAL KNOWLEDGE OF GMA ADEQUATE?

There is more disagreement than agreement on this question across the set of articles in this special issue. Some authors even questioned the existence of anything meaningful in GMA, beyond an abstraction. For example, Goldstein et al. (this issue) noted that there is no agreement on the definition and measurement of GMA and quoted Gould (1994) for how reification of an abstract concept results in absurdities. This lack of agreement is also echoed by Sternberg and Hedlund (this issue). Reeve and Hakel (this issue) observed that all constructs are abstractions, and, thus, the criticism of GMA as only an abstraction is misguided. They also noted that discussions of "Psychometric g" involve a redundant adjective; all factors are psychometric. The use of terms such as "psychometric g" or "the so-called general factor," as though GMA is a statistical phenomenon with no practical importance, makes sense for scoring debate points, but is scientifically meaningless.

In criticizing the theoretical basis of GMA for predicting job performance, Goldstein et al. (this issue) charges that the finding of a simple, linear relation generalizable across jobs between GMA and job performance has (a) stunted the development of the theoretical basis and refinements of other predictors and influences on job performance and (b) restricted us to linear models to the exclusion of an exploration of nonlinear, configural models. Dissenting, Gottfredson (this issue) points out that theoretical obtuseness should not be used to discredit or confuse the literature to satisfy socially motivated goals. She points out the presence of a general factor in many test batteries, in different samples, in different cultures, and so forth. Gottfredson notes the lawful relation between the *g*-loadedness of tests and their validity in predicting complex task performance. The case of personality assessment is invoked to stress that the presence of a general factor cannot be an artifact of factor analyses.

What process mechanisms explain the effects of cognitive ability on performance and other work outcomes? Ree and Carretta (this issue) as well as

Schmidt (this issue) summarize existing path models developed to test process mechanisms of the effects of GMA on work outcomes. Although the research is not discussed by the contributing authors, we would also note the several laboratory experiments that clarify the role of GMA in work-relevant task performance. There may be more explorations of the theoretical rationale for GMA-outcome relations in IWO psychology than other fields.

Ree and Carretta (this issue) summarize the several psychological and biological correlates of GMA. Reeve and Hakel (this issue) state that knowledge of how GMA is important for a wide variety of life outcomes is sparse. After reviewing the material from behavior genetics, they proposed some potential reasons for racial group differences. Future research should explore other mechanisms to better understand how and why GMA differs across individuals and across groups defined by ethnic origin. Advances in biological sciences, and behavior genetics in particular, may advance our understanding of GMA, notably on these questions.

Several authors advise the incorporation of findings from other fields in the quest to better understand the theoretical basis of GMA. Schmidt (this issue) suggest that a course in differential psychology must be compulsory for all industrial–organizational psychologists. Tenopyr (this issue) stress the need to assimilate the recent advances in cognitive and developmental psychology to understand the development of GMA in individuals. She called for linking the study of individual differences and organizational research. Tenopyr also mentioned the research on person–environment fit and the concept of "typical intellectual engagement" proposed by Ackerman and his colleagues (e.g., Ackerman 1989) as potentially fruitful avenues for future research to explore. In Tenopyr's view, although current theories of GMA enjoy some support, (a) they can all be criticized, and (b) the possibility of a viable new theory cannot be dismissed.

Finally, we should stress that the theoretical knowledge of GMA is much more than what we have for other predictors. Murphy (1996) characterizes this state of affairs noting that research on predictors in personnel selection can be classified as either those exploring cognition or those exploring variables other than cognition. In fact, Goldstein et al. (this issue) use this massive (compared to other predictors) theoretical database on GMA as evidence of how GMA has stunted development and refinement of other predictors. It appears that there are important unresolved questions about GMA, but we know more about GMA than any other construct used as a predictor in personnel selection.

IS THERE PROMISE IN NEW METHODS OF TESTING FOR GMA?

Goldstein et al. (this issue) suggest that the low cost of paper-and-pencil tests has killed the demand for other types of testing media. Some efforts have been made to

assess GMA with different technologies. To reduce group differences on paper-and-pencil GMA tests, a strategy that seems to be gaining in popularity appears to be changes in test medium (Sackett, Schmitt, Ellingson, & Kabin, 2001). Computerized and video-based assessments (cf. Chan & Schmitt, 1997) are some examples. Salgado and Anderson (this issue) discussed a virtual reality technology used in Europe for selection. Sternberg and Hedlund (this issue) listed alternative approaches to assessing practical intelligence. If the use of a different medium reduces adverse impact without reducing validity for a criterion, then the new method of testing is preferable.

In pursuing alternative media and test format changes, care must be taken not to alter the construct being assessed. "Format changes" may rarely be truly format changes alone. Scores on new formats of assessing GMA may reflect individual differences in GMA plus individual differences in responding to the new medium of testing. If the latter source of variance also reduces group differences but is unrelated to the criterion of interest, then changing the medium of assessment may be beneficial. Murphy (this issue) suggests that this might be one approach to try to balance efficiency and (group) equity. However, Schmidt (this issue) as well as Kehoe (this issue) correctly argue that, given the lack of differential prediction for paper-and-pencil GMA tests, any reduction in group differences resulting from changes in testing medium will result in differential prediction. Also, in assessing the relative reduction of group differences resulting from format changes, measurement error must also be taken into account (see the previous section for a discussion). What might be perceived as a medium effect may actually be a reliability effect. Emphasizing this point, Gottfredson (this issue) and Schmidt (this issue) suggest that most of the recent attempts to reduce group differences by changes to GMA test format may have been due to changes in the level of measurement error in GMA assessments.

The next few decades will be interesting, as new tools based on physiological, biological, and even genetic markers are identified for GMA. Whether these potentially more invasive assessments are developed and become available for use depends on how society decides to balance the privacy rights of individuals against the needs of organizations.

WHAT IS THE CURRENT STATUS OF NON-GMA PREDICTORS AS SUBSTITUTES OR SUPPLEMENTS TO GMA?

The search for alternative predictors to GMA has already had a long history (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). The first issue where we note differences across our 11 sets of contributors is whether we should consider the different predictors as substitutes to GMA or as supplements to GMA. Outtz (this

issue) argue that because GMA produces 3 to 5 times the adverse impact as other equally valid predictors, it is indefensible to use GMA as a major component in a selection system. Murphy (this issue) states that we cannot defend the use of GMA if alternate predictors of equal validities can be found. On the other hand, Schmidt (this issue) noted that alternatives are not substitutes, but supplements. This is also echoed by Gottfredson (this issue), who classified predictors as can do, will do, and have done predictors. GMA tests capture a can do construct, personality measures reflect will do constructs, and work sample tests are prototypical examples of have done predictors. Gottfredson noted that GMA predicts task performance, whereas personality variables could better predict contextual performance (Borman & Motowidlo, 1997). However, she stresses that no organization would forego selecting for core technical performance. Thus, alternative predictors can at best serve as supplements, but never as substitutes for GMA.

This view is also echoed by Goldstein et al. (this issue), whose emphasis is on increasing the percentage variance explained in job performance by using more complex models and incorporating non-GMA measures into selection systems. Similar views are also expressed by Sternberg and Hedlund (this issue). We should note that, although reductions in group differences are possible by combining GMA and noncognitive measures with negligible group differences such as integrity tests (Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Schmidt, 1993), this strategy cannot completely erase the group differences (Sackett & Ellingson, 1997).

The alternate predictors frequently cited in the literature include personality variables, job knowledge measures, and work sample tests. Tenopyr (this issue) notes that, despite the recent favorable reviews of personality assessment for predicting job performance, the actual validities are just as low as those reported by Guion and Gottier (1965). She also raises issues of faking and the need to develop applicant level norms. Not all personality constructs and measures have equal validity in predicting overall job performance. As we have noted elsewhere, compound personality measures such as integrity tests (Ones et al., 1993) and Criterion-focused Occupational Personality Scales (Ones & Viswesvaran, 2001) produce validities in upper .30s and lower .40s for overall job performance. Such criterion-related validities are certainly superior to those reported for the Big Five personality dimensions (Barrick & Mount, 1991; Hurtz & Donovan, 2000). As for social desirability influences on the validity of personality measures, extensive research has shown that faking among job applicants does not destroy criterion-related validity (Hough, 1998; Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996). However, applicant norms may be essential to appropriately interpret the meaning of scores in personnel selection settings.

Research on job knowledge and work samples has concentrated on several predictors such as job knowledge tests, tacit knowledge tests, situational judgment tests, and so forth. Gottfredson (this issue) note that these "have done" predictors cannot

be used with untrained applicants, for most entry-level positions, or with inexperienced individuals (a point also made by Schmidt, this issue). Furthermore, job knowledge and job sample performance are partly consequences of GMA.

Sternberg and Hedlund (this issue) promote the use of tacit knowledge as an alternate predictor. However, many authors (Gottfredson; Ree, & Carretta; Reeve & Hakel; Tenopyr; all this issue) express concern about the construct definition of tacit knowledge. Sternberg and Hedlund note that tacit knowledge is a specific kind of procedural knowledge. There is also added clarity that Sternberg and colleagues consider practical intelligence a determinant of tacit knowledge. Sternberg and Hedlund argue that the empirical evidence to date suggests that tacit knowledge is independent of GMA. However, correlations between the measures have been examined in very range restricted samples (e.g., Yale undergraduates; see Kuncel et al., 1998, and Kuncel, Hezlett, & Ones, 2001, for detailed discussions of this point). More important, if tacit knowledge is indeed a type of procedural knowledge, it may be misleading to measure it among those who have had no opportunity to gain this knowledge (e.g., college students). Among art majors, it would be relatively easy to show that physics knowledge does not correlate with GMA, but this does not mean that acquiring physics knowledge is unrelated to one's intelligence!

As Goldstein et al. (this issue) note in their article, often the distinction between cognitive abilities and noncognitive predictors is not as clear cut as we would like it to be. The distinction is fuzzy at the boundaries. This raises questions as to whether the predictive validity of alternate predictors is partly due to GMA. Tenopyr (this issue) wonder whether some of the incremental validity of interviews and biodata reported in some studies are due to interviews and biodata measuring cognitive abilities, not measured by the GMA test used in those studies. For alternate predictors such as personality and psychomotor ability tests, the assessment process may invoke GMA. Goldstein et al. note how cognitive demands can exist even in personality assessments, and a similar point is made by Ree and Carretta (this issue) with respect to assessment of psychomotor ability. Finally, Goldstein et al. note that working memory capacity may influence both performance in GMA tests as well as on alternate predictors.

In general, the theoretical foundation of alternate predictors is nowhere near that accumulated for GMA (Murphy, 1996). In evaluating alternate predictors, one should apply the same standards to these as those applied to GMA tests. Unfortunately, most of the extant literature fails to do this. It is also important to note that, if alternatives are used as substitutes rather than as supplements, by forgoing use of GMA tests we may be missing an important predictor of performance (particularly core task performance).

There are two issues that future research should address with respect to the development of alternate predictors. First, there seems to be some confusion between constructs and methods (Hough & Oswald, 2000; Schmitt & Chan, 1998). It is not clear

what one refers to when the validity of (or group differences in) interviews, biodata, assessment centers, and so forth, are discussed. Interviews, assessment centers, and biodata can be used to measure all sorts of different constructs from social skills to abstract mathematical problem-solving skills. Furthermore, an interview used as a first selection hurdle is likely to differ in content and form from an interview used as a final step in the selection process.

A second issue is the lack of empirical data on the predictor intercorrelations. Both to build comprehensive theories of work performance as well as to design personnel selection systems to maximize validity and reduce group differences, we need to accurately and precisely estimate predictor intercorrelations. Schmitt, Rogers, Chan, Sheppard, and Jennings (1997) compiled a meta-analytically derived matrix (cf. Viswesvaran & Ones, 1995) to estimate these intercorrelations. Bobko, Roth, and Potosky (1999) cautioned us about the methodological choices involved in putting together such a matrix. More important, it is clear that there is very sparse empirical evidence to guide us here.

A related issue in assessing predictor intercorrelations is the type of selection system under consideration. The implicit assumption in using these predictor intercorrelation matrices to investigate the validity and adverse impact of different combinations (Sackett et al., 2001; Schmitt et al., 1997) is that the selection systems are fully compensatory. When predictors are used in different stages, then there is a need to account for different levels of range restriction. That is, in considering the intercorrelation of one predictor with other predictors, we need to consider the order in which the predictors will be used. Thus, if there are six predictors, then we may need to construct $2^6 - 1 = 63$ intercorrelation matrices for the different combinations. Ultimately, the needs of organizations might dictate how predictor incorrelation matrices are put together and used.

A final crucial point to note in considering other predictors is to be aware that equal validity may not imply equal value. This view underlies even the Uniform Guidelines on Employee Selection Procedures (U.S. Equal Employment Opportunity Commission, 1978) in that the guidelines suggest that when there are equally valid predictors, those with less adverse impact should be employed. This assumption that equal validity implies equal value is also reflected in utility models. However, as Kehoe (this issue) notes

> A correlation coefficient of .25 between a personality composite and a measure of overall performance does not necessarily reflect the same value for an organization as a .25 correlation between GMA and overall performance. The people chosen based on the personality measure would not be the same, and they would not have the same attribute profile, as the people chosen based on GMA. The two equal correlations would reflect different construct mappings of predictors onto the criterion, although the variance accounted for is the same. Operationally, the organization might experience the differences in a variety of tangible ways. For example, the "personal-

ity" employees might achieve their overall performance by being relatively more dependable, persistent, attentive, helpful, and so on. The "cognitive" employees might achieve their overall performance by being relatively more accurate, faster, effective problem solvers, and the like.

This view of equal validity being different from equal value should be kept in mind when considering the multiattribute utility models discussed by Murphy (this issue). Murphy states that, logically, decision makers should prefer selection system A over system B if (a) A is higher in both equity and efficiency, (b) A equals B in efficiency but A is greater than B in terms of equity, and (c) A equals B in equity but A is greater than B in efficiency. The key point to note here is that efficiency cannot be defined solely in terms of validity, and equity has to be clearly taken into consideration.

CONCLUSION

The set of 11 articles in this special issue are thought-provoking and raise several important points. These articles summarize what psychologists have discovered about GMA thus far. Of interest, contributors also diverge on the future importance of GMA for job performance. Predictions are diametrically opposing: The role of GMA in the sphere of work may (a) increase (Kehoe, this issue) and (b) decrease (Goldstein et al., this issue). We tend to believe that, with increasing complexity in our environments, the former is more likely than the latter. Increasing technological sophistication of the workplace as well as globalization of economic activities (see Salgado and Anderson, this issue) are likely to contribute to the increase in the importance of GMA. Hopefully, this special issue of *Human Performance* will serve as a summary for practitioners and guide future research. We recognize that goals of any society are not determined solely by research findings or the scientific literature. Research and literature can only point out whether the goals set by society are likely to succeed or to fail. It is our hope that a clearer understanding of GMA and its role in IWO psychology will maximize GMA's usefulness for applied purposes and will increase decision quality when setting public policy.

ACKNOWLEDGMENT

Both authors contributed equally to this article; the order of authorship is arbitrary.

REFERENCES

Ackerman, P.L., (1989). Within task intercorrelations of skilled performance: Implications for predicting individual differences. Journal of Applied Psychology, 74, 360–364.

Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology, 77*, 836–874.

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

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. *Personnel Psychology, 52*, 561–589.

Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587–605.

Borman, W. C. (1978). Exploring the upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology, 63*, 135–144.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64*, 410–412.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco: Jossey-Bass.

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance, 10*, 99–109.

Brand, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 251–265). New York: Falmer.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 42*, 143–159.

Ford, J. K., Kraiger, K., & Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. *Psychological Bulletin, 99*, 330–337.

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694–734.

Gilliland, S. W. (1995). Fairness from the applicant's perspective: Reactions to employee selection procedures. *International Journal of Selection and Assessment, 3*, 11–19.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132.

Gould, S.J., (1994). Curveball. *The New Yorker,* November 28.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135–164.

Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology, 39*, 811–826.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hough, L.M., (1998). Effects of distortion in personality measurement and suggested palliatives. *Human Performance, 4,* 209–244.

Hough, L. M., & Oswald, F. L. (2000). Personnel selection: Looking toward the future—Remembering the past. *Annual Review of Psychology, 51*, 631–664.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.

Hunter, J. E., & Schmidt, F. L. (1977). A critical analysis of the statistical and ethical definitions of test fairness. *Psychological Bulletin*, *83*, 1053–1071.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law, 2,* 447–472.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology, 70*, 56–65.

Kuncel, N. R., Campbell, J. P., & Ones, D. S. (1998). The validity of the Graduate Record Exam: Estimated or tacitly known? *American Psychologist, 53*, 567–568.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162–181.

Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin, 87*, 72–107.

Lubinski, D., & Humphreys, L. G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*, *24*, 159–201.

Murphy, K. R. (1996). Individual differences and behavior in organizations: Much more than *g.* In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 3–30). San Francisco: Jossey-Bass.

Murphy, K. R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage Publications.

Murphy, K. R., & DeShon, R. P. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53.* 873–900.

Ones, D.S., Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. International Journal of Selection and Assessment, 9, 31–*39.*

Ones, D. S., & Viswesvaran, C. (1996). Bandwidth-fidelity dilemma in personality measurement for personnel selection. *Journal of Organizational Behavior, 17*, 209–226.

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

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

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

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97*, 307–315.

Pulakos, E. D., Oppler, S. H., White, L. A., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology, 74*, 770–780.

Ree, M.J., & Earles, J.A. (1991). Predicting training success: Not much more than g. *Personnel Psychology, 44,* 321–332.

Rotundo, M., & Sackett, P. R. (1999). Effect of rater race on conclusions regarding differential prediction in cognitive ability tests. *Journal of Applied Psychology, 84*, 815–822.

Rynes, S. L. (1993). Who's selecting whom? Effects of selection practices on applicant attitudes and behavior. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 240–274). San Francisco: Jossey-Bass.

Sackett, P. R., & DuBois, C. L. Z. (1991). Rater–ratee race effects on performance evaluation: Challenging meta-analytic conclusions. *Journal of Applied Psychology, 76*, 873–877.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology, 50*, 707–721.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*, 549–572.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (in press). High-states testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist*.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, *36*, 1128–1137.

Schmidt, F. L., & Hunter, J. E. (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.

Schmidt, F. L., Viswesvaran, C., & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology, 53*, 501–912.

Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Thousand Oaks, CA: Sage.

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.

Schuler, H., Farr, J. L., & Smith, M. (1993). (Eds.). *Personnel selection and assessment: Individual and organizational perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Silva, J. M., & Jacobs, R. R. (1993). Performance as a function of increased minority hiring. *Journal of Applied Psychology, 78*, 591–601.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology, 46*, 49–76.

U.S. Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*, 38290–38315.

Viswesvaran, C. (1993). *Modeling job performance: Is there a general factor*? Unpublished doctoral dissertation, University of Iowa, Iowa City.

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

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8 , 216*–227.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557–560.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Uses, consequences, and controversies. Part 1: Report of the committee.* Washington, DC: National Academy Press.