Journal of Personality and Social Psychology
2004, Vol. 86, No. 1, 148–161

0022-3514/04/$12.00 DOI: 10.1037/0022-3514.86.1.148

Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All?

Nathan R. Kuncel and Sarah A. Hezlett
University of Illinois at Urbana–Champaign

Deniz S. Ones
University of Minnesota, Twin Cities Campus

This meta-analysis addresses the question of whether 1 general cognitive ability measure developed for predicting academic performance is valid for predicting performance in both educational and work domains. The validity of the Miller Analogies Test (MAT; W. S. Miller, 1960) for predicting 18 academic and work-related criteria was examined. MAT correlations with other cognitive tests (e.g., Raven's Matrices [J. C. Raven, 1965]; Graduate Record Examinations) also were meta-analyzed. The results indicate that the abilities measured by the MAT are shared with other cognitive ability instruments and that these abilities are generalizably valid predictors of academic and vocational criteria, as well as evaluations of career potential and creativity. These findings contradict the notion that intelligence at work is wholly different from intelligence at school, extending the voluminous literature that supports the broad importance of general cognitive ability (*g*).

Many laypeople, as well as social scientists, subscribe to the belief that the abilities required for success in the real world differ substantially from what is needed to achieve success in the classroom. Yet, this belief is not empirically or theoretically supported. A century of scientific research has shown that general cognitive ability, or *g*, predicts a broad spectrum of important life outcomes, behaviors, and performances. These include academic achievement, health-related behaviors, social outcomes, job performance, and creativity, among many others (see Brand, 1987; Gottfredson, 1997; Jensen, 1998; Lubinski, 2000; Ree & Caretta, 2002; Schmidt, 2002, for reviews of variables that display important relations with cognitive ability). A particularly powerful demonstration of the influence of *g* comes from Jencks et al. (1979) who showed that even with background and socioeconomic status (SES) controlled, cognitive ability measured at adolescence predicted occupational attainment. Cognitive ability "is to psychology as carbon is to chemistry" (Brand, 1987, p. 257) because it truly impacts virtually all aspects of our lives.

How is it that many people believe that the abilities required for success in the real world differ substantially from what is needed to achieve success in the classroom? Perhaps the fact that tests and measures are often developed for particular settings (e.g., educational vs. occupational) has perpetuated this myth. The main purpose of the current study is to evaluate whether a single test of cognitive ability that was developed for use in educational settings is predictive of behaviors, performances, and outcomes in both educational and occupational settings. We first conduct a series of meta-analyses to establish that the Miller Analogies Test (MAT; Miller, 1960) assesses cognitive ability. We then report meta-analyses examining the validity of the MAT for predicting multiple criteria in academic and work settings, including evaluations of career potential and creativity. The results address the theoretical question of whether a single cognitive ability measure is valid for predicting important criteria across domains. In this article, general cognitive ability and *g* are defined as the underlying trait that leads to the well-documented positive intercorrelation observed between measures of cognitive behaviors. The phenomenon of *g* has been shown to have important, domain-general relationships with knowledge, learning, and information processing, and the general thesis of this article is that tests of general cognitive ability or *g* are predictive of success in academic and work settings, regardless of the setting for which they were developed.

Although our thesis and findings may surprise some readers, it was our a priori expectation that the MAT would be a valid predictor of a wide range of academic and work criteria, as well as creativity and career potential. Our prediction was based on the enormous literature that unequivocally demonstrates the existence of a general factor of cognitive ability and its broad importance as a predictor of numerous life outcomes (for reviews, see, Brand, 1987; Gottfredson, 2002). Therefore, this study builds on and contributes to the substantial body of research already supporting the nomological network in which the construct of *g* is embedded.

Nathan R. Kuncel, Department of Psychology, University of Illinois at Urbana–Champaign; Sarah A. Hezlett, Department of Human Resource Education, University of Illinois at Urbana–Champaign; Deniz S. Ones, Department of Psychology, University of Minnesota, Twin Cities Campus.

Nathan R. Kuncel gratefully acknowledges the National Science Foundation for indirect support for this project through a graduate research fellowship and the University of Minnesota for indirect support of this project through an Eva O. Miller fellowship. We thank John P. Campbell, Marcus Crede, Mark L. Davison, Ates Haner, Lloyd Humphreys, and Frank L. Schmidt for their helpful comments and suggestions. We also thank Barton Adams, Brian Griepentrog, Yoshani Keiski, Jeanette Shelton, David Sowinski, and John Morton for their assistance in gathering the articles summarized in this meta-analysis and Jennifer Vannelli and Lisa L. Thomas for their assistance with manuscript preparation.

Correspondence concerning this article should be addressed to Nathan R. Kuncel, Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820. E-mail: nkuncel@uiuc.edu

To clarify how and why a test of *g* can be predictive of multiple criteria in different domains of life (e.g., school and work), we provide an overview of this nomological network. We first focus on the nature and structure of *g* as defined by research on the relations among cognitive ability measures. We then discuss the relationship between *g* and learning. This relationship provides the basis for a discussion of theories that specify how *g* has a positive relationship with job performance. Last, we describe how the nature of the tasks that constitute academic and job "work" leads us to expect that *g* would predict performance in both settings.

The Nature of *g*: Evidence for Generality Across Cognitive Ability Tests

In 1904, following a critical review of prior research on intelligence and an analysis of data he collected from school children, Spearman concluded "that all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (1904, p. 284). Scores on all tests of cognitive ability were proposed to be a function of two components: a general component, *g*, and a specific component, *s* (Spearman, 1927). The specific component is unique to a test or limited set of tests and cannot be measured without simultaneously measuring *g* (Spearman, 1937).

Despite considerable evidence supporting its existence, the concept of a pervasive general factor has remained controversial since its introduction. Many later models of intelligence are based on the idea that there are a number of independent specific abilities (Ree & Caretta, 1998). For example, on the basis of the first major application of factor analysis, Thurstone (1938) proposed that there are seven primary cognitive abilities. Later, Guilford (1959) argued that over 100 distinct abilities exist in his structure of intellect (SOI) model.

The existence of truly separate abilities is not supported by the data. Spearman (1939) himself demonstrated that a reanalysis of Thurstone's (1938) data yielded a general factor, an empirical finding later acknowledged by Thurstone (1947). A reanalysis of Guilford's data by Alliger (1988) yielded a robust average correlation of .45 among SOI measures. Studies of the contemporary triarchic theory (Sternberg, 1985) find a general positive correlation among measures of the three intelligences (e.g., Sternberg, Castejón, Prieto, Hautakami, & Grigorenko, 2001). These and many other studies undermine the concept that there are independent broad cognitive abilities. Scores on tests of cognitive abilities covary because of a general factor, *g*.

On the basis of considerable empirical evidence examining the structure of cognitive abilities, a hierarchical model ultimately replaced Spearman's (1904) two-component conceptualization (Carroll, 1993; Ree & Caretta, 1998). A hierarchical model is widely accepted as the best means of representing the communality among measures of cognitive abilities, with *g* being the highest order factor (Carroll, 1993). Descending lower in the hierarchy, there is a generally agreed on set of increasingly more specific abilities (e.g., Carroll, 1993; Drasgow, 2002; Gustaffson, 1984; Vernon, 1961). A pervasive general factor across tests developed for different purposes (e.g., assessing academic achievement vs. practical intelligence) and for use in different domains (e.g., educational vs. occupational) forms the basis of cross-situation and cross-domain generalizability expectations for the predictive value of general cognitive ability tests.

Why does *g* demonstrate an impressive array of external correlates that are commensurate with its dominant presence among cognitive ability measures? We first discuss the relationships among *g* and three highly related areas: learning, training success, and skill acquisition. The strong relationship between *g* and the acquisition of knowledge and skill (learning) sets the stage for a discussion of the powerful relationship between *g* and job performance. Several theories of the relationship between *g* and job performance are grounded in the relationship between *g* and learning.

g and Learning: The Theoretical Basis for the Cross-Domain Predictive Value of *g*

The existence of a relationship between *g* and learning is not without dissension. For example, Gardner (1983) has written that general cognitive ability, "reveals little about an individual's potential for further growth" (p. 18) and "foretells little of success in later life" (p. 3). Statements such as these ignore a mountain of data. Some dramatic evidence of the relationship between *g* and growth, learning, and success comes from a longitudinal study by Lubinski, Webb, Morelock, and Benbow (2001). They found that a group of profoundly gifted students who obtained exceptional SAT scores before the age of 13, had achieved, 10 years later, a long list of impressive accomplishments including numerous scientific publications, many exceptional original pieces of writing and art, and several inventions. Certainly, development from a 13-year-old to a published scientist or a 19-year-old graduate from the Massachusetts Institute of Technology reflects exceptional growth and success.

Less dramatic, but no less informative about the relationship between *g* and learning, are the numerous studies that demonstrate the strong positive correlation among *g* and educational success[1] (e.g., Kuncel, Hezlett, & Ones, 2001; Linn & Hastings, 1984), skill acquisition (e.g., Ackerman, 1987, 1992; Lohman, 1999), and job training success (for reviews, see Ree & Caretta, 1998; Schmidt, 2002). These studies represent just some of the many meta-analyses and comprehensive reviews summarizing the literature that support the relationship between *g* and what are effectively different operationalizations of learning. On the basis of this literature, defining *g*, in part, as an ability or capacity to learn and acquire new knowledge and skill (e.g., Cattell, 1971; Schmidt, 2002; Snyderman & Rothman, 1987) is very appropriate.

The relationship between *g* and job training success warrants extra attention because of its connection to the real world. Time and again, research has documented large positive predictive validities between cognitive ability tests and performance in job training programs. Primary studies have been compiled in large-scale meta-analyses examining the validity of cognitive ability measures for training success in a number of specific jobs or job

[1] Before dismissing educational success studies as "just grades," the reader should note that many studies have found strong predictive validity for important non-GPA criteria (e.g., Kuncel, Hezlett, & Ones, 2001) and that GPAs are far from trivial as they predict salary level (Roth & Clarke, 1998) and job performance (Roth, BeVier, Switzer, & Schippman, 1996).

categories (Hirsh, Northrup, & Schmidt, 1986; Lilienthal & Pearlman, 1983; Schmidt, Hunter, Pearlman, & Shane, 1979). For example, Thorndike (1985) reported that an ability composite weighted to approximate *g* correlated, on average, .65 with final grades attained at 35 Army Technical Schools. Overall, there is considerable evidence that *g* is a universal predictor of job training success (for reviews, see Hunter, 1980, 1986; Hunter & Hunter, 1984; Jensen, 1986; Ree & Earles, 1991). More specific cognitive abilities account for only a limited amount of variance in training criteria above and beyond *g* (Ree & Earles, 1991; Thorndike, 1985).

The training literature also helps explain why *g* is an important predictor of job performance and success in other life areas that require learning and performing cognitively loaded tasks. Using meta-analysis to evaluate alternate models of training, Colquitt, LePine, and Noe (2000) concluded that cognitive ability influenced the acquisition of declarative knowledge (job knowledge) and procedural knowledge (job skills).[2] Skill acquisition mediated the relationship between cognitive ability and executing the new skills in the work environment (i.e., transfer of training), which in turn was related to actual job performance. Theories and research on the determinants of job performance also offer support for the relationship between *g* and the acquisition of job-relevant knowledge and skill.

g and Job Performance: Generalizable Relationships and Empirically Grounded Theories

Prior to 1977, applied psychologists and employers presumed that the abilities required for job performance were job specific and differed substantially from job to job. However, Schmidt and Hunter (1977) demonstrated that variability across jobs and settings largely were due to sampling error and other statistical artifacts. Subsequent meta-analyses showed that cognitive ability test validities are generalizably valid across jobs, situations, and settings (see Schmidt, 2002; Schmidt & Hunter, 1998; Schmidt, Ones, & Hunter, 1992, for reviews). There is now abundant evidence supporting the validity of cognitive ability measures for predicting work-related criteria in both civilian and military organizations (e.g., Hunter, 1983, 1986; Hunter & Hunter, 1984; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Pearlman, Schmidt, & Hunter, 1980; Reeve & Hakel, 2002; Thorndike, 1985). In fact, "*g* can be said to be the most powerful predictor of overall job performance" (Gottfredson, 1997, p. 83).

Current theories explaining the universal relationship between *g* and job performance emphasize a related set of mediating variables. For example, a theory of work performance proposed by Campbell and his colleagues (Campbell, 1990; Campbell, Gasser, & Oswald, 1996) outlines eight dimensions of job performance and their determinants. McCloy, Campbell, and Cudeck (1994) provided empirical support for this model, demonstrating that performance on the job is a direct function of declarative knowledge, procedural knowledge, and motivation. General cognitive ability was shown to be indirectly related to job performance through its influence on declarative knowledge, procedural knowledge, and motivation.

These results are similar to the results from other researchers (Borman, Hanson, Oppler, Pulakos, & White, 1993; Schmidt & Hunter, 1993; Schmidt, Hunter, & Outerbridge, 1986) who proposed and empirically demonstrated that job knowledge (declarative knowledge) and skill (procedural knowledge) mediate the relationship between *g* and job performance. Their results suggest that job knowledge is a direct determinant of job performance, whereas *g*'s influence on job performance is primarily indirect. General cognitive ability predicts job performance across jobs and settings primarily because it predicts learning and acquisition of job knowledge.

Job knowledge requirements and the complexity of jobs tend to go hand in hand. Not all job knowledge or job tasks are equally difficult or complex. The literature has demonstrated that the validity of *g* for predicting job performance is moderated by the complexity of the job. Research by Hunter (1983) found that the validity of *g* ranges from a low of .23 for low-complexity jobs (e.g., shrimp picker, cannery worker) up to a correlation of .58 for high-complexity jobs (e.g., retail food manager, fish and game warden). In her review of this literature, Gottfredson (1997) made the important observation that jobs do not need to be more "academic" for there to be a stronger relationship with *g*. The job only needs to comprise more cognitively complex tasks.

It is likely that complexity moderates the relationship between *g* and job performance for two reasons. First, more complex jobs require the worker to acquire a greater amount of complex knowledge (e.g., tank turbine engine schematics vs. cash register layout). Second, more complex jobs require more difficult information processing (e.g., synthesizing financial data vs. adding up a restaurant bill). In general, the ubiquitous finding that *g* predicts performance in all jobs and its validities typically exceed those for other single traits (for reviews, see Gottfredson, 1997; Ree & Caretta, 1998; Schmidt, 2002; Schmidt & Hunter, 1998) stems from the fact that those who score high on *g* are far more likely to have acquired considerable declarative and procedural knowledge, in addition to being able to quickly acquire (both in formal training programs and on the job) additional, job-specific declarative and procedural knowledge and process complex information.[3]

The Nature of "Work" Tasks: Parallels Between Educational and Occupational Work

The research literature that demonstrates a consistent positive relationship between *g* and academic as well as work performance may surprise those who see the "work" in both contexts as being fundamentally different. Some have argued that academic tasks are

[2] Declarative knowledge (DK) refers to the "knowledge of facts, rules, principles, and procedures. Specifically, DK represents the ability to state the facts, rules, principles, or procedures that are a prerequisite for successful task performance (Anderson, 1985; Kanfer & Ackerman, 1990)" (McCloy et al., 1994, p. 493). Procedural knowledge is the "capability attained when DK (knowing what to do) has been successfully combined with knowing how and being able to perform a task" (McCloy et al., 1994, p. 493).

[3] Note that these findings are consistent with the theory of fluid and crystallized intelligence, which suggests that specific knowledge is the result of an investment of general cognitive ability into the development of more specific abilities or knowledge (Cattell, 1971), as well as the more elaborate theory proposed by Ackerman (1996), which also includes the influences of personality and interest variables on the acquisition and development of knowledge and skill.

different from practical or real-world tasks (e.g., Sternberg & Wagner, 1993). Academic tasks from this perspective are said to be well defined, have only a single correct answer, and be self-contained, among other things. We argue that the accuracy of this perspective is limited to only a subset of examinations, as many examinations are more complex, requiring tasks such as analysis and synthesis (Bloom, Hastings, & Madaus, 1971). More important, this description does not begin to do justice to the complex behaviors students engage in before they sit down to complete an examination or to do other academic assignments, such as writing term papers or making oral presentations. Like work tasks, many academic tasks are complex and ill defined. They lack a single right answer and often require students to obtain additional information and generate creative solutions. Effectively limiting a definition of academic performance to behavior at the time of the final examination results in a misleading picture of the nature of academic performance and the individual difference determinants that are likely to be related to performance inside and outside of the classroom. We propose that the nature and determinants of academic performance are similar, although not fully identical, to the nature and determinants of job performance.

Work settings emphasize the application of previously acquired declarative and procedural knowledge with a lesser, but still critical, emphasis on acquiring new declarative and procedural knowledge. In an academic setting, a greater emphasis is placed on directly demonstrating that declarative knowledge has been recently acquired. For example, course examinations, papers, comprehensive examinations, oral examinations, and dissertation defenses are focused on testing an individual's current level of knowledge in a specific area. In heavily cumulative disciplines (e.g., mathematics, chemistry), performance is also partially a function of previously acquired, discipline-specific declarative and procedural knowledge. For example, prior knowledge and skill solving mathematical problems influences the acquisition of new mathematical knowledge. However, academic performance is not just the production of recently acquired knowledge.

Academic performance in the classroom is the end product of many other behaviors. For example, obtaining a good grade after answering examination items is the result of effective performance studying, managing goal conflicts, coordinating work with classmates, seeking additional information, negotiating with peers and faculty, avoiding counterproductive behaviors (e.g., drugs and alcohol), handling finances, and structuring effective communications (e.g., Kuncel, Campbell, Hezlett, & Ones, 2001; Reilly, 1976). Each of these is likely to be partially determined by declarative and procedural knowledge, such as specific study skills, writing skills, planning skills, and team performance skills. The extent to which students have mastered these skills varies across individuals and is partially a function of *g* as well as other individual differences.

In summary, performance in both academic and work settings is a direct function of learned declarative and procedural knowledge. Performance in the workplace is directly determined by the motivated application of declarative and procedural knowledge, with a lesser emphasis on acquiring additional knowledge and skill. Performance in an academic classroom setting is determined by the direct demonstration of declarative and procedural knowledge after having engaged in many other complex and ill-defined tasks; that is, the knowledge was recently acquired through a number of different and complex tasks that occur both within and outside of the classroom.

Therefore, although the academic setting places a greater emphasis on the acquisition of knowledge, performance in both settings should be and is predicted by *g*. Both situations involve learning. Both situations contain complex or practical tasks. Finally, performance in both situations is partially determined by previously acquired levels of knowledge and skill. General cognitive ability is related to all three of these. As a result, the same cognitive ability measure should be a valid predictor of performance in both settings even if that measure was originally developed for use in academic admissions.

Overview of Current Study

Despite this substantial body of theory and research, the applicability of abilities needed in academic settings to work settings continues to be questioned. The vast theoretical and empirical literature reviewed above speaks loud and clear: Any test that assesses *g* will have predictive value in both settings, demonstrating cross-domain generalizability. The purpose of our study is to demonstrate this finding using the same cognitive ability test. Unfortunately, most instruments are not used in both domains. A notable exception is the Miller Analogies Test (MAT; Miller, 1960; The Psychological Corporation, 1994). We meta-analytically examined the extent to which (a) the MAT assesses cognitive ability, (b) the MAT predicts important behaviors and outcomes in educational settings, and (c) the MAT predicts job performance and other important criteria in work settings. As such, this research offers a direct test of whether the abilities related to performance in academic settings overlap with those predicting performance in work settings.

The Miller Analogies Test

The MAT has been used for admissions decisions into graduate schools as well as hiring/promotion decisions for moderate- to high-complexity jobs in industry (under the name MAT as well as the Advanced Personnel Test). Developed by W. S. Miller at the University of Minnesota and in use since 1926, the 100-item MAT is composed entirely of analogy items (Miller, 1960; The Psychological Corporation, 1994). The items in the MAT differ from many analogy questions that are almost entirely dependent on a participant's ability to reason with vocabulary knowledge. Instead, the MAT is composed of analogies that require knowledge of many domains including sciences, vocabulary, literature, arts, and history. The MAT is a timed test (although most test takers finish all of the items), with test takers given 50 min to complete the 100 items. Retesting is permitted with one of several alternate forms. There is no penalty for guessing.

Previous studies of the MAT have often arrived at different conclusions about its predictive validity (e.g., Marascuilo & Gill, 1967; Nagi, 1975; Perney, 1994; Watters & Paterson, 1953). Most of these studies have ignored the effects of statistical artifacts on the results, including differences in unreliability, range restriction, and sampling error. As a result, some authors have concluded that, even in academic settings, the validity of the MAT is likely to be heavily moderated by situational factors. The present study also tests this claim.

Hypotheses

An important step in our research is linking scores on the MAT to other measures of cognitive ability. If the MAT not only correlates very highly with the cognitive ability measures that have strong validity evidence for predicting academic performance but also correlates very highly with the cognitive ability measures that are predictive of work performance, it would be reasonable to conclude that the MAT and related tests are valid in academic and work settings.

In developing hypotheses about how the MAT correlates with other cognitive tests, it is important to consider what it measures. On the basis of the hierarchical models discussed earlier, scores on the MAT are a function of *g*, verbal reasoning abilities, and abilities specific to the test. The influence, or weight, of *g* should be high. Spearman (1923, 1927) noted that analogy tests were correlated with other measures known to capture *g*. In describing his three basic principles of cognition (apprehension of experience, eduction of relations, and eduction of correlates), he relied heavily on analogy problems for illustration. Analogical reasoning involved all three principles of cognition, making analogies a superb measure of *g* (Bejar, Chaffin, & Embretson, 1991). Therefore, we expect scores on the MAT will be strongly correlated with tests of general cognitive ability and reasoning.

Verbal ability also appears to be needed to solve the kinds of analogy problems that constitute the MAT. Although the role played by vocabulary level in the difficulty of verbal analogy items varies by the nature of the analogy involved (Bejar et al., 1991), knowledge of the meaning of words is clearly needed in answering this kind of question. Therefore, because they share both *g* and verbal ability in common, we predict that the MAT will have exceptionally strong correlations with other tests of verbal ability.

In contrast, we anticipate that the MAT will have strong, but lower, correlations with tests of mathematical ability. Although the MAT has some problems involving math, a smaller proportion of the items are related to mathematical ability than verbal ability. In other words, mathematical ability has less weight than verbal ability in this test. To recapitulate, we predict the MAT to correlate near unity with verbal tests. We expect the next largest correlation to be with the general cognitive ability and reasoning tests, followed by correlations with quantitative tests.

On the basis of the vast literature yielding information on the relations among *g*, declarative and procedural knowledge, job performance, and performance in educational and training settings, our expectation is that a single, *g*-loaded test, such as the MAT, will be a valid predictor of performance in both work and academic settings. In each domain, certain indexes of performance tend to be heavily used (namely, grade point average and overall performance ratings). However, performance in neither setting is unidimensional (Campbell, 1990; Campbell et al., 1996; Enright & Gitomer, 1989; Kuncel, 2003; Reilly, 1976; Viswesvaran & Ones, 2000). Consistent with the idea that job and task complexity moderate the relationship between *g* and job performance, we predict scores on the MAT will be better predictors of some criteria than others.

In general, we expect the MAT to have correlations with academic criteria that are nearly identical to those of the Graduate Record Examination—Verbal test (GRE–V; Briel, O'Neill, & Scheuneman, 1993) and a pattern of correlations with work criteria that is consistent with the literature on the consistently positive correlation between cognitive ability and job performance. Lower correlations are anticipated between the MAT and tasks in both academic and work domains that are heavily motivation-loaded. Substantial proportions of the variance in these criteria are likely to be determined by differences in the direction, level of intensity, and duration of individuals' effort.

More specifically, we expect the MAT to have moderate positive relationships with a number of academic criteria, including 1st-year grade point average (GPA), graduate GPA (GGPA), and faculty ratings of performance. Given the verbal content of the MAT and verbal nature of comprehensive examinations, we expect large, positive correlations between the MAT and comprehensive examination scores. We anticipate a small but positive correlation between the MAT and research productivity because of the low base rate for students engaging in publishing and the fact that many graduate students are not interested in preparing for or pursuing a research career. We expect degree attainment to be largely determined by motivation and, therefore, predict the MAT to have a positive but small correlation with degree attainment. Within the group of students who ultimately finish their degree program, speed of completion seems to be almost entirely driven by volitional choice and key situational factors, such as financial resources and departmental characteristics (Baird, 1990; Bowen & Rudenstine, 1992). Therefore, we predict that the MAT will have a zero correlation with time to complete. Number of courses completed appears to be largely determined by personality traits related to compliance or motivation. Most students who enter graduate school have the ability to, at a minimum, complete a course (with, perhaps, less than stellar performance). We expect a small but positive correlation between the MAT and number of courses completed. Overall, those criteria that are strongly related to acquiring domain specific knowledge (e.g., grades, faculty ratings, comprehensive examinations) will tend to have the strongest relationships, those that are more distal products of knowledge as well as interest (e.g., research accomplishment, degree attainment) will have more modest relationships, and those that are almost purely motivationally determined (e.g., number of courses, time to complete) will have the lowest correlations.

In our search of the literature, we located several studies reporting relationships between graduate students' MAT scores and their performance on worklike criteria. These criteria, which include internship/practicum performance ratings, student-teaching performance ratings, and performance on a counseling work sample, fall in between academic and job criteria. In preparation for their future roles, students are asked to complete simulations or execute tasks in a work setting that are highly similar to those employees perform on the job. These criteria embody the transition between school and work. Because of the substantial literature indicating that *g* predicts performance in both academic and work domains, we predict a moderate correlation between the MAT and internship/practicum performance ratings and performance on a counseling work sample. Smaller correlations are expected between the MAT and student-teaching performance ratings because of the high probability of construct-irrelevant contamination in the criterion. Raters had very limited opportunities to observe the teacher, and the behaviors to be rated were often ill specified or primarily interpersonal in nature. Thus, a measure of *g* will account for only

a small proportion of the variance in the student-teaching performance ratings.

Ratings of creativity is a fourth criterion we examined that incorporates elements of academia and work. Ratings of creativity were a faculty member's or work supervisor's evaluation of a person's creativity or potential for creative work. The majority (60%) of the studies involved evaluations of student creativity rather than employee creativity. We believe that the ability determinants of creative work are mainly composed of *g*, related specific abilities, and acquired domain specific knowledge. We expected moderate correlations between the MAT and ratings of creativity.

Similarly, a number of studies reported correlations between the MAT and ratings of potential for either students or employees. Ratings of potential for future performance were for either success in a job (e.g., counseling) or an aspect of a job (e.g., research). Ratings of potential for work in counseling constituted 55% of the studies for the potential criterion. These studies were included in the overall potential analyses, as well as being examined separately. On the basis of previous research in assessment centers and potential ratings (Gaugler, Rosenthal, Thornton, & Bentson, 1987), we hypothesized moderate correlations between the MAT and potential ratings.

The work-setting criteria include job performance and membership in professional organizations. Consistent with prior research on general cognitive ability and job performance, we expect a moderate and positive correlation between the MAT and job performance. Performance as an educational administrator was analyzed separately from the other job performance studies because many of the ratings were for what we felt were nonwork characteristics and behaviors (e.g., global self-confidence, emotional stability). This mixture of work and nonwork ratings will probably result in positive but smaller correlations between the MAT and educational administrator performance.

Membership in a professional organization was simply whether the person, postgraduation, was a member in a professional society (e.g., American Psychological Association). Note that this criterion was constrained to membership and did not address being elected as a fellow or an officer of an organization. We expected a small but positive correlation between the MAT and membership in a professional organization on the basis of the expectation that individuals who were more effective performers are more likely to join professional organizations.

Method

To quantitatively aggregate results across previous studies of the MAT, we used Hunter and Schmidt's (1990) psychometric meta-analytic method. As has been documented by its use within the field of industrial–organizational psychology, meta-analysis is a particularly powerful method for clarifying research in an area. By statistically aggregating research on a topic, it increases the amount of information that can be brought to bear on a single question. To summarize the literature, we began by computing the average, sample-size/weighted correlation across all studies (r_{obs}). For each estimate, the corresponding standard deviation of the observed correlations was also calculated (SD_{obs}).

Some meta-analytic procedures, such as those developed by Hunter and Schmidt (1990), also help clarify research by effectively addressing the effects of undesirable study characteristics that influence the magnitude and distribution of observed correlations. These statistical artifacts include sampling error, restriction of range, and measurement unreliability.

For example, the attenuating effect of studies that use samples that are highly restricted in range on the predictor test can be reduced or eliminated (e.g., students from an elite school). This is important because the process of using a predictor (e.g., the MAT) to select a group of new graduate students or workers results in a sample that is less variable on the predictor. As a result, correlations involving the predictor (assuming they are not zero) will be attenuated (Thorndike, 1949). This downward bias results in an underestimate of the actual relationship of the predictor with other variables for the population of interest. In this study, all potential applicants to a graduate program or job were considered to be the population of interest. Meta-analysis also can be used to account for the reality that studies differ in terms of the reliability of the measures that they use. The unreliability of measures artificially lowers their observed correlation.

Correcting r_{obs} and SD_{obs} for range restriction and measurement error yields more accurate estimates of the relationship between two variables and permits evaluation of whether the variability in observed correlations is due to systematic artifactual biases or reflects the existence of substantive moderators. Furthermore, correcting SD_{obs} for the sometimes large differences in sample sizes across studies yields a more accurate estimate of whether the differences observed in the literature are merely the result of sampling error.

In examining the relationship of the MAT with other cognitive ability measures, as well as with academic and work-related criteria, we were interested in understanding the true relationship between the variables. In essence, the goal was to create the best estimate of the population correlation in the absence of all sampling error, restriction of range, and measurement error. To create this estimate (ρ), all possible corrections for statistical artifacts were made (i.e., corrections for range restriction, along with corrections for the unreliability of both variables).

An estimate of interest to those who are using a test for admissions or hiring is the operational validity (r_{op}). Operational validity refers to the test-criterion correlation coefficient that has been corrected for unreliability in the criterion but not in the predictor. Because selection or admissions decisions are made with an imperfectly reliable measure, predictor unreliability corrections are not applied when estimating the operational validity of the test. Operational validity answers the applied question of what would be gained if the predictor were used for hiring or admissions purposes.[4] For each MAT–criteria relationship, we also estimated operational validity.

Corrections for range restriction and unreliability were also applied in computing variability estimates across the correlations included in each meta-analysis. The standard deviation of observed correlations corrected for statistical artifacts is the residual standard deviation (SD_{res}). The standard deviation of the true score validities (SD_ρ) describes the standard deviation associated with the true validity, after variability that is due to sampling error, unreliability in the predictor, unreliability in the criterion, and range restriction have been removed. The magnitude of SD_ρ is an indicator for the presence of moderators. Smaller values suggest that other variables are unlikely to substantially moderate the validity of the MAT. If all or a major portion of the observed variance in a correlation is due to statistical artifacts, one can conclude that the relationship is constant or nearly so.

The SD_ρ is also used to compute lower 90% credibility interval, which is used as an indicator of the likelihood that the true relationship generalizes across situations. If the lower 90% credibility value is greater than zero, one can conclude that the presence of a relationship can be generalized to new situations (Hunter & Schmidt, 1990). In our meta-analysis, if the 90% credibility value is greater than zero, but there is variance in the correlations after corrections, it can be concluded that the relationships of

[4] Because the MAT is a highly reliable test, the applied validity and generalizability of the test differs little from its theoretical validity when predictor unreliability is corrected.

the MAT with other variables are positive across situations, although the actual magnitude may vary somewhat across settings. However, the remaining variability may also be due to uncorrected statistical artifacts, other methodological differences, and unidentified moderators.

In making corrections, because not all studies included in our database reported the necessary measurement error and range-restriction information, we relied on the extant research literature to construct appropriate unreliability and range-restriction distributions. These artifact distributions are described in the Appendix.

We gathered studies involving prediction of graduate school performance and job performance by the MAT from several sources. To identify relevant research *PsycLIT* (1887–2001) and ERIC (1966–2001) searches were combined with a search of *Dissertation Abstracts International* (1861–2000). The citation lists within all articles, dissertations, and technical reports were also examined to identify additional relevant studies. Unreported effect sizes were computed from available information when possible.

In articles with sample overlaps, the larger or more complete data were included in the meta-analysis, and the matching or overlapping samples were excluded. When a single study reported two or more correlations between the MAT and measures of the same general criterion, the correlations were averaged to create an overall estimate.

Sorting of data for inclusion in each meta-analysis was based on the consensus of the three authors. The final database for the criterion-related meta-analyses included 163 independent samples from 127 studies, yielding 229 correlations across 20,352 subjects. No single analysis included more than one correlation from the same sample so independence was not violated. The list of studies contributing data to our study may be obtained by contacting the first author. Data were analyzed with interactive meta-analysis procedures (Schmidt, Gast-Rosenberg, & Hunter, 1980), using a program developed by Schmidt, Hunter, and Viswesvaran (1998), with improvements that increase accuracy over Hunter and Schmidt's (1990) original method.[5]

Results

We first present the correlations between the MAT and other cognitive ability measures. Then, we describe the results for the eight criteria in graduate school settings. Third, we discuss findings on the value of the MAT for predicting the performance of graduate students on work-related tasks, as well as for predicting ratings of creativity and potential for students and employees. Last, we report results for the four criteria from the work setting.[6]

Correlations With Other Ability Measures

Sufficient studies have reported correlations of scores on the MAT with scores on the GRE–V and the GRE—Quantitative test (GRE–Q) to permit separate analyses. The remaining tests were classified into three categories: primarily verbal tests, primarily mathematical tests, and tests assessing *g* and reasoning. Examples of tests assessing verbal ability were the MCAT Verbal and Cooperative Reading Test—Total Score. Tests evaluating mathematical ability included the MCAT Quantitative and Doppelt Mathematical Reasoning. Examples of tests included in the general cognitive ability and reasoning category included the Raven Progressive Matrices (Raven, 1965), Army Alpha (Yerkes, 1921), Watson Glaser (Watson & Glaser, 1980), and Wechsler-Bellevue (Wechsler, 1946).

The results in Table 1 indicate that the MAT is very strongly associated with measures of verbal ability. The validity of the MAT for predicting GRE–V and other verbal tests were equally strong at .88 (k = 15, N = 8,328 and k = 23, N = 3,614, respectively) with relatively low values of SD_ρ (.06 and .08, respectively), which indicate there is little true variability across studies.

The relationships between the MAT and tests of mathematical ability are smaller but still large, with true correlations of the MAT with the GRE–Q and other math tests equaling .57 (k = 15, N = 7,055) and .68 (k = 18, N = 2,874), respectively. Small values of SD_ρ for GRE–Q (.10) and other mathematical ability measures (.05) indicate there is little room for moderators to operate. Our results indicate that the MAT is also strongly related to performance on other general cognitive ability and reasoning tests (ρ = .75, k = 15, N = 1,753). The associated SD_ρ is slightly larger (.15).

Criterion-Related Validities

Results for all academic criteria are shown in Table 2. Validity for GGPA was moderately large (k = 70, N = 11,368) with a true-score correlation (ρ) of .39 and a small standard deviation of true validity of .09. The ρ for 1st-year GPA was .41, which was slightly larger than ρ for GGPA (k = 34, N = 2,999) with, also, a larger SD_ρ of .15. Faculty ratings (k = 25, N = 1,909) were professors' ratings of graduate student performance. All of the ratings are separate from internship/teaching ratings, were focused on performance in graduate school, and did not include ratings of behaviors from outside of graduate school (e.g., empathy, life balance). Validities for this criterion are also presented in Table 2 and were moderately large, with a ρ of .37. The estimate for SD_ρ was 0, resulting in a 90% credibility interval that did not include zero. These findings indicate that the MAT is a generalizably valid predictor of faculty ratings and that including the somewhat heterogeneous mixture of faculty ratings was not unreasonable, as the variability in the results was fully accounted for by statistical artifacts.

Comprehensive examination score results were based on studies that reported a correlation with either comprehensive examination or preliminary examination scores. The validity of the MAT was largest for this criterion, with a ρ of .58 and a small (.03) SD_ρ, and was based on a more modest sample size (k = 10, N = 987) than the GGPA, 1st-year GPA, and faculty rating criteria.

Research productivity is an important goal of many graduate students and programs. We found four studies that investigated how well the MAT predicts research productivity (N = 314). The results suggest that the MAT has a small, positive relationship with research productivity (ρ = .19). The SD_ρ of zero indicates that statistical artifacts account for all variance across the four studies. All four of the studies included in this analysis were doctoral programs emphasizing research, and not professional programs

[5] These refinements included use of the mean observed correlation in the formula for sampling error variance and the use of a nonlinear range restriction formula to estimate the standard deviation of corrected validities (Law, Schmidt, & Hunter, 1994a, 1994b).

[6] For consistent rules of thumb for the interpretation of effect size magnitude, see Lubinski and Humphreys (1997). The reader should note that Cohen (1992) provided r and d rules of thumb that differ in magnitude.

Table 1
Meta-Analysis of Miller Analogies Test (MAT) Correlations With Other Ability Tests

Test	N	k	r_{obs}	SD_{obs}	SD_{res}	ρ	SD_ρ	90% cred.
Graduate Record Examination—Verbal	8,328	15	.70	.08	.05	**.88**	.06	.78–.98
Other verbal ability measures	3,614	23	.67	.10	.06	**.88**	.08	.75–1.01
Graduate Record Examination—Quantitative	7,055	15	.42	.10	.07	**.57**	.10	.41–.73
Other math ability measures	2,874	18	.50	.09	.03	**.68**	.05	.60–.76
General ability and reasoning measures	1,753	15	.56	.14	.11	**.75**	.15	.50–1.00

Note. For all analyses, ρ has been corrected for MAT reliability, other test reliability, and restriction of range. Boldface type highlights the most critical information. N = sample size, k = number of studies; r_{obs} = sample size weighted average correlation; SD_{obs} = standard deviation of observed correlations; SD_{res} = residual standard deviation; ρ = estimated true score correlation; SD_ρ = standard deviation of true score correlations; 90% cred. = 90% credibility interval.

designed to train practitioners, and may not generalize to master's or professional degree programs.

Degree attainment was generally defined as studies in which the MAT was used to predict taking the graduate degree versus failing to take the degree. A couple of studies were included in which the contrast was made between dropping out, being in program, and graduating from the program. For this criterion, ρ was positive (ρ = .21, k = 20, N = 3,963). The SD_ρ was larger for this criterion (.24) than other criteria. This, combined with the relatively small operational validity, resulted in a lower 90% credibility interval that included zero. These results suggest that the MAT has a generally positive modest relationship with degree attainment but that the validity is likely to be moderated by other factors.

Five studies were located that examined the relationship between MAT scores and how long it took students to finish their degrees (N = 1,700). The true score correlation of .35 indicates that the MAT moderately predicts time to finish degree. The relatively small SD_ρ of .11 suggests that any third variable is likely to have a limited moderating effect. The number of courses/credits completed appears to be effectively unrelated to MAT scores (ρ = –.06) and was based on a small sample size and number of studies (k = 3, N = 179). This criterion is only a function of the number of courses the student actually completed, not the ratio of attempted to completed.

Three criteria were identified that essentially represent practice or simulated job performance: internship/practicum ratings, student-teaching performance ratings, and counseling work sample performance. The findings for these criteria are in Table 3. Internship and practicum ratings had a small, positive relationship with the MAT (ρ = .22), whereas student-teaching performance had a near zero negative association with the MAT (ρ = –.04). In both cases, SD_ρ equaled zero. For internship/practicum ratings and student-teaching performance ratings, the number of studies and students were limited (ks = 4 and 5 and Ns = 300 and 444, respectively), and therefore we cannot reach unequivocal conclusions regarding these two criteria. Moderate validity was obtained for predicting counseling work sample performance (ρ = .27, k = 5, N = 114, SD_ρ = 0).

Results for ratings of creativity and potential also are shown in Table 3. Creativity ratings were made by faculty or work supervisors, and creativity rating validities were not included in the analyses for any other criteria. The true score correlation for predicting these ratings was a moderate .36 (k = 6, N = 1,104, SD_ρ = 0). The MAT is a strong predictor of the counseling potential ratings subsample (ρ = .49, k = 6, N = 192) and a moderate predictor for the overall potential ratings (ρ = .37, k = 11, N = 494) analysis. For both criteria the SD_ρs of zero indicate the validity of the MAT generalizes across situations.

The MAT has been used to predict several work and professional criteria, including job performance and membership in professional organizations. The results for these criteria are shown in Table 4. On the basis of seven studies, the true score correlation for predicting job performance was estimated to be .41 (N = 598, SD_ρ = .12). For two jobs, there were sufficient studies to warrant

Table 2
Meta-Analysis of Miller Analogies Test (MAT) Validity for Academic Criteria

Criterion	N	k	r_{obs}	SD_{obs}	SD_{res}	r_{op}	ρ	SD_ρ	90% cred.
Graduate grade point average	11,368	70	.27	.10	.06	.36	**.39**	.09	.24–.54
1st-year graduate grade point average	2,999	34	.29	.15	.11	.38	**.41**	.15	.16–.66
Faculty ratings	1,909	25	.25	.11	.00	.34	**.37**	.00	.37–.37
Comprehensive examination scores[a]	987	10	.47	.10	.03	.54	**.58**	.03	.53–.63
Research productivity[a,b]	314	4	.13	.07	.00	.17	**.19**	.00	.19–.19
Degree attainment[a,b]	3,963	20	.15	.19	.17	.20	**.21**	.24	−.18–.60
Time to finish degree[a,b]	1,700	5	.25	.10	.08	.32	**.35**	.11	.17–.53
Number of courses/credits completed[a]	179	3	−.05	.05	.00	−.06	**−.06**	.00	−.06–−.06

Note. Boldface type highlights the most critical information. N = sample size; k = number of studies; r_{obs} = sample size weighted average correlation; SD_{obs} = standard deviation of observed correlations; SD_{res} = residual standard deviation; r_{op} = estimated validity for applied use of the MAT (i.e., operational validity); ρ = estimated true score validity (ρ has been corrected for MAT unreliability, criterion unreliability [where possible], and restriction of range); SD_ρ = standard deviation of true score correlations; 90% cred. = 90% credibility interval.
[a] Not corrected for criterion unreliability. [b] Corrected using the more restricted MAT range-restriction distribution (see Appendix).

Table 3
Meta-Analysis of Miller Analogies Test (MAT) Validities for Transitional Criteria, Creativity, and Potential

Variable	N	k	r_{obs}	SD_{obs}	SD_{res}	r_{op}	ρ	SD_{ρ}	90% cred.
Internship/practicum ratings[a]	300	4	.13	.10	.00	.20	**.22**	.00	.22–.22
Student-teaching performance ratings[a]	444	5	−.02	.09	.00	−.04	**−.04**	.00	−.04–−.04
Counseling work sample performance[b]	114	5	.18	.14	.00	.25	**.27**	.00	.27–.27
Ratings of creativity[c]	1,104	6	.25	.06	.00	.34	**.36**	.00	.36–.36
Potential ratings[d]	494	11	.24	.13	.00	.34	**.37**	.00	.37–.37
Counseling potential ratings[d]	192	6	.32	.10	.00	.45	**.49**	.00	.49–.49

Note. Boldface type highlights the most critical information. N = sample size; k = number of studies; r_{obs} = sample size weighted average correlation; SD_{obs} = standard deviation of observed correlations; SD_{res} = residual standard deviation; r_{op} = estimated validity for applied use of the MAT (i.e., operational validity); ρ = estimated true score validity (ρ has been corrected for MAT unreliability, criterion unreliability [where possible], and restriction of range); SD_{ρ} = standard deviation of true score correlations; 90% cred. = 90% credibility interval.
[a] Corrected using job performance reliability estimate. [b] Corrected for unreliability using a distribution of work sample reliability. [c] Corrected using reliability distribution of potential ratings. [d] Corrected using reliability distribution of creativity ratings.

separate analyses. Counseling performance was strongly predicted (ρ = .51, k = 2, N = 92, SD_{ρ} = 0). Note that these data were nested within the overall job performance analysis. The MAT was less strongly associated with educational administration performance (ρ = .27, k = 10, N = 225, SD_{ρ} = .15). These data are independent from the job performance analyses. The results suggest that the MAT is a generalizably valid predictor of work performance criteria.

Last, three studies were found that examined the relationship between MAT scores and membership in a professional organization. The true score correlation of .27 indicates that individuals who perform better on the test are more likely to be members of professional organizations (k = 3, N = 278, SD_{ρ} = 0). However, the results are based on a limited number of samples and people.

Discussion

The MAT is a valid predictor of several aspects of graduate student performance as well as measures of job performance, potential, and creativity. The validity was at least as high for work criteria as for school criteria. The MAT was a valid predictor of seven of the eight measures of graduate student performance (average ρ = .32), five of the six school-to-work transition performance criteria (average ρ = .29), and all four of the work performance criteria (average ρ = .37). Consistent with the voluminous existing literature, our results provide compelling evidence that an ability instrument that was developed to aid educational admissions decisions is also valid for predicting job performance. These findings, in combination with the strong correlations between the MAT and other cognitive ability tests from educational and work settings, provide direct evidence that g is related to success in multiple domains.

At the same time, the pattern of correlation between the MAT and individual criteria also supports the importance of more specific abilities. The MAT was an exceptionally good predictor of comprehensive examinations. Given that the correlations among the MAT and other ability measures indicate that the MAT measures verbal ability in addition to g, it is reasonable to conclude that some of the strength of its relationship with comprehensive examination scores is due to its verbal component. The GRE–V and MAT have nearly identical large correlations with comprehensive examinations (and other criteria) whereas the GRE–Q is a weaker predictor of comprehensive examinations (Kuncel et al., 2001).

The argument that intelligence at work is completely different from intelligence at school is not supported by our results. Instead the results are consistent with the large body of research that suggests that performance is a function of the motivated application of acquired declarative (i.e., job knowledge) and procedural knowledge (i.e., skill; Borman et al., 1993; McCloy et al., 1994; Schmidt et al., 1986). Acquiring declarative and procedural knowledge is strongly related to g.

The reader should note that the estimates obtained in this study are underestimates of the relationship between g and performance

Table 4
Meta-Analysis of Miller Analogies Test (MAT) Validities for Work Criterion Variables

Variable	N	k	r_{obs}	SD_{obs}	SD_{res}	r_{op}	ρ	SD_{ρ}	90% cred.
Job performance[a,b]	598	7	.26	.14	.07	.38	**.41**	.12	.21–.61
Counseling performance[a,b]	92	2	.33	.05	.00	.47	**.51**	.00	.51–.51
Educational administration performance[a,b]	225	10	.15	.23	.10	.25	**.27**	.15	.02–.53
Membership in a professional organization[b,c]	278	3	.19	.19	.00	.25	**.27**	.00	.27–.27

Note. Boldface type highlights the most critical information. N = sample size; k = number of studies; r_{obs} = sample size weighted average correlation; SD_{obs} = standard deviation of observed correlations; SD_{res} = residual standard deviation; r_{op} = estimated validity for applied use of the MAT (i.e., operational validity); ρ = estimated true score validity (ρ has been corrected for MAT unreliability, criterion unreliability [where possible], and restriction of range); SD_{ρ} = standard deviation of true score correlations; 90% cred. = 90% credibility interval.
[a] Corrected using job performance reliability estimate. [b] Corrected using the more restricted MAT range-restriction distribution (see Appendix). [c] Not corrected for criterion unreliability.

if the full range of talent in the U.S. population had been considered. We defined our population of interest as applicants to graduate schools or more technical jobs and used normative data to make appropriate range-restriction corrections to these more homogeneous groups. Therefore, our results address the relationship between the MAT and performance in various domains given the realities of the educational and credentialing systems in the United States. In the theoretical case, where people with very low ability applied to doctoral-level history programs or for professional engineering positions that currently require a postsecondary degree, the predictive power of general cognitive ability would probably be even greater for academic and work performance than our findings indicate.

The results from this study also indicate that the abilities measured by the MAT and other cognitive ability tests are related to evaluations of creativity and potential. These findings are consistent with other research findings that indicate that exceptional scores on educational admission tests are associated with exceptional academic and creative accomplishments (Lubinski et al., 2001). Selecting students or workers on the basis of cognitive ability results in scholars and employees who are creative and high potential.

Although most of our hypotheses were supported, there were a few surprises that should be discussed. First, the correlation between the MAT and work performance, although a strong positive correlation, is somewhat lower than has been observed in similar meta-analyses of cognitive ability tests for predicting job performance (see Hunter & Hunter, 1984). We see three explanations (beyond sampling error) for this difference. First, our restriction of range corrections is smaller than corrections from the work performance and graduate performance domains. Our conservative corrections may have resulted in an underestimate. Second, some of the job performance measures used in this study are clearly incomplete and construct deficient. Third, a superior measure of *g* would be the composite of several different ability measures. Therefore, the MAT is a somewhat constant deficient measure of general cognitive ability.

A second unexpected result was that the MAT had a positive correlation with time to finish the graduate degree. Our best (ad hoc) explanation is that more able students are likely to spend time in graduate school doing nondegree work (e.g., research) that may keep them from finishing as fast as other students. Overall, if the only myopic goal for a program is rapid completion, then one could negatively weight the MAT. If the goal is to train students who finish at higher rates, acquire field-specific knowledge, are well regarded by the faculty, produce research, and are viewed as creative and having potential, one should positively weight the MAT. The best choice seems crystal clear to us.

The effectively zero correlation between the MAT and student-teaching ratings was also inconsistent with our hypotheses. We believe that this result is due to the quality and focus of student-teaching ratings. Limited opportunities to observe and poorly specified criteria for good teaching may account for the zero correlation. Alternatively, it may be that the results are simply due to sampling error. Being knowledgeable in basic math, science, and English is clearly important for good teaching. Given the fact that the MAT correlates with the acquisition and demonstration of knowledge and skill (e.g., GPA, comprehensive examinations, counseling work samples), it is unlikely that good teaching (properly measured) is unrelated to the abilities measured by the MAT.

In general, the MAT was weakly correlated with more motivationally determined criteria such as degree attainment. Before the reader dismisses some of the smaller validities as too small to be of use, one should note that the percentage increase of correct decisions that can be obtained with a small correlation can be substantial, particularly when an organization can be highly selective (Taylor & Russell, 1939). The MAT would still have utility for increasing the number of students who finish a degree program. Nonetheless, there is considerable room for improving the prediction of criteria such as degree attainment, and previous research has shown that noncognitive predictors can be particularly effective for primarily noncognitive criteria (e.g., McHenry et al., 1990). Finishing a PhD program requires a great deal of goal-directed behavior, and measures of interests and personality are likely to be useful (and superior) predictors of more motivationally determined criteria as well as providing incremental validity for more cognitively determined criteria. This illustrates a broader point that specific dimensions of performance will be best predicted by different combinations of *g*, interests, personality, and other individual differences. There is no denying the important influence of other traits on behavior, and we hope the surge of interest in personality within the workplace (e.g., Hough & Ones, 2001; Roberts & Hogan, 2001) spills over into research on academic performance. However, across many aspects of life and particularly for learning and job performance, we like the succinct summary from Drasgow (2002, p. 126) that "*g* is key".

We acknowledge that the analyses are based on sample sizes that range from rather limited to very large and that the clustering of criterion measures was partially a subjective, nonetheless consensual, decision among the three coauthors. We felt that our clustering maximized the use and interpretability of the available information. However, alternative clusterings are certainly worth considering and debating. It is also important to note that some possibility of criterion contamination exists in these studies. Faculty, and to a lesser extent, employers may be aware of individuals' MAT scores. We believe the influences are likely to be negligible and note that the results are completely consistent with research examining ability measures using completely predictive designs (see Schmidt et al., 1992, for a discussion). Overall, we found considerable evidence that the validity of the MAT generalizes across academic and work settings and feel that the overall pattern of test intercorrelation results and predictive validity results creates a compelling case for the importance of *g* across academic and work settings.

The MAT was shown to fit within the established ability literature as a measure of both *g* and verbal ability. Our overall hypothesis that the same cognitive ability instrument would predict performance in academic and work settings was supported. The abilities measured by the MAT predict more than just grades. They predict other academic criteria as well as a variety of professional and work performance criteria, including evaluations of creativity and potential. Furthermore, our results suggest that the lower correlations and much of the variability observed in previous research are likely to have been the result of range restriction, criterion unreliability, and sampling error. Arguments that cognitive abilities measured by educational admissions tests are not related to behaviors and outcomes beyond grades early in a degree

program are contradicted by the empirical data presented in this study. Cognitive ability tests, such as the MAT, are predictive of success in academic and work settings, regardless of the setting for which they were developed.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77,* 598–614.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22,* 227–257.

Alliger, G. M. (1988). Do zero correlations really exist among measures of different intellectual abilities? *Educational & Psychological Measurement, 48,* 275–280.

Baird, L. L. (1990). Disciplines and doctorates: The relationships between program characteristics and the duration of doctoral study. *Research in Higher Education, 31,* 369–385.

Barritt, L. S. (1966). Note: The consistency of first-semester college grade point average. *Journal of Educational Measurement, 3,* 261–262.

Bejar, I. I., Chaffin, R., & Embretson, S. (1991). *Cognitive and psychometric analysis of analogical problem solving.* New York: Springer-Verlag.

Bendig, A. W. (1953). The reliability of letter grades. *Educational and Psychological Measurement, 13,* 311–321.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisor experience in supervisor performance. *Journal of Applied Psychology, 78,* 443–449.

Bowen, W. G., & Rudenstine, N. L. (1992). *In pursuit of the Ph.D.* Princeton, NJ: Princeton University Press.

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.

Briel, J. B., O'Neill, K., & Scheuneman, J. D. (Eds.). (1993). *GRE technical manual.* Princeton, NJ: Educational Testing Service.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol. 1* (2nd ed., pp. 687–732). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P., Gasser, M. B., & Oswald, F. L. (1996). The substantive nature of job performance variability. In K. R. Murphy (Ed.), *Individual differences and behavior in organizations* (pp. 258–299). San Francisco: Jossey-Bass.

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

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton Mifflin.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Colquitt, J. A., LePine, J. A., & Noe, R. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85,* 678–707.

Drasgow, F. (2002). Intelligence and the workplace. In W. C. Borman, D. R. Ilgen, R. J. Klimoski (Eds.), *Handbook of psychology, Volume 12, industrial and organizational psychology* (pp. 107–130). New York: Wiley.

Enright, M. K., & Gitomer, D. (1989). *Toward a description of successful graduate students* (GRE Board Report No. 85-17R). Princeton, NJ: Educational Testing Service.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences* (2nd ed.). New York: Basic Books.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72,* 493–511.

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

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist, 14,* 469–479.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8,* 179–203.

Hirsh, H. R., Northrup, L., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39,* 399–420.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity, and use of personality variables in industrial, work, and organizational psychology. In N. Anderson, D. S. Ones, H. Sinangil, & C. Viswesvaran (Eds.) *Handbook of industrial, work, and organizational psychology: Vol. 1* (pp. 233–277). London, UK: Sage.

Hunter, J. E. (1980). *Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB).* Washington, DC: U.S. Employment Service.

Hunter, J. E. (1983). *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (GATB)* (Test Research Rep. No. 45). Washington, DC: U.S. Employment Service, U.S. Department of Labor.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340–362.

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. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G. et al. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books, Inc.

Jensen, A. R. (1986). g: Artifact or reality? *Journal of Vocational Behavior, 29,* 301–331.

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

Kuncel, N. R. (2003). *The prediction and structure of academic performance.* Unpublished doctoral thesis, University of Minnesota, Minneapolis.

Kuncel, N. R., Campbell, J. P., Hezlett, S. A., & Ones, D. S. (2001, April). Performance in college: The criterion problem. In D. S. Ones & S. A. Hezlett (Chairs), *Predicting performance: The interface of I/O psychology and educational research.* Symposium presented at the annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

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.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994a). Non-linearity of range corrections in meta-analysis: Test of an improved procedure. *Journal of Applied Psychology, 79,* 425–438.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994b). A test of two refinements in procedures for meta-analysis. *Journal of Applied Psychology, 79,* 978–986.

Lilienthal, R. A., & Pearlman, K. (1983). *The validity of federal selection*

tests for aid/technician in the health, science, and engineering fields. Washington, DC: U.S. Office of Personnel Management, Office of Personnel Research and Development.

Linn, R. L., & Hastings, C. N. (1984). A meta-analysis of the validity of predictors of performance in law school. *Journal of Educational Measurement, 21,* 245–259.

Lohman, D. F. (1999). Minding our p's and q's: On finding relationships between learning and intelligence. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences* (pp. 55–76). Washington, DC: American Psychological Association.

Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points." *Annual Review of Psychology, 51,* 405–444.

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

Lubinski, D., Webb, R. M., Morelock, M. J, & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86,* 718–729.

Marascuilo, L. A., & Gill, G. (1967). Measurable differences between successful and unsuccessful doctoral students in education. *California Journal of Educational Research, 18,* 65–70.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79,* 493–505.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335–354.

Miller, W. S. (1960). *Technical manual for the Miller Analogies Test.* New York: The Psychological Corporation.

Nagi, J. L. (1975). Predictive validity of the Graduate Record Examination and the Miller Analogies Tests. *Educational and Psychological Measurement, 35,* 471–472.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training criteria in clerical occupations. *Journal of Applied Psychology, 65,* 373–407.

Perney, J. (1994, April). *Using a writing sample to predict success in masters programs in education.* Paper presented at the annual meeting of the American Education Research Association, New York. (ERIC Document Reproduction Service No. ED395999).

The Psychological Corporation. (1994). *Technical manual for the Miller Analogies Test: A guide to interpretation.* San Antonio, TX: Author.

The Psychological Corporation. (2001). *Miller Analogies Test: Candidate information booklet.* San Antonio, TX: Harcourt Assessment Company.

Raven, J. C. (1965). *Progressive matrices.* New York: The Psychological Corporation.

Ree, M. J., & Caretta, T. R. (1998). General cognitive ability and occupational performance. *International Review of Industrial and Organizational Psychology, 13,* 159–184.

Ree, M. J., & Carretta, T. R. (2002). g2K. *Human Performance, 15,* 3–23.

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

Reeve, C. L., & Hakel, M. D. (2002). Asking the right questions about *g. Human Performance, 15,* 47–74.

Reilly, R. R. (1976). Factors involved in graduate student performance. *American Education Research Journal, 13,* 125–138.

Reilly, R. R., & Warech, M. A. (1993). The validity and fairness of alternatives to cognitive tests. In L. C. Wing & B. R. Cifford (Eds.), *Policy issues in employment testing* (pp. 131–224). Boston: Kluwer.

Roberts, B. W., & Hogan, R. (Eds.). (2001). *Personality psychology in the workplace.* Washington, DC: American Psychological Association.

Roth, P. L., Be Vier, C. A., Switzer, F. S., & Schippman, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81,* 548–556.

Roth, P. L., & Clarke, R. L. (1998). Meta-analyzing the relation between grades and salary. *Journal of Vocational Behavior, 53,* 386–400.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15,* 187–210.

Schmidt, F. L., Gast-Rosenberg, I. F., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65,* 643–661.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529–540.

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability and job knowledge. *Current Directions in Psychological Science, 2,* 8–9.

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., Hunter, J. E., & Outerbridge, A. H. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432–439.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization model. *Personnel Psychology, 32,* 257–281.

Schmidt, F. L., Hunter, J. E., & Viswesvaran, C. (1998). [Meta-analysis MS-Basic software]. Unpublished software.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42,* 137–144.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15,* 201–293.

Spearman, C. (1923). *Nature of 'intelligence' and principles of cognition.* London: MacMillan.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Spearman, C. (1937). *Psychology down the ages, Volume II.* London: Macmillan.

Spearman, C. (1939). Thurstone's work re-worked. *Journal of Educational Psychology, 30,* 1–16.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.

Sternberg, R. J., Castejón, J., Prieto, M. D., Hautakami, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test (multiple-choice items) in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment, 17,* 1–16.

Sternberg, R. J., & Wagner, R. K. (1993). The *g*-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science, 2,* 1–5.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical validity of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23,* 565–578.

Thorndike, R. L. (1949). *Personnel selection.* New York: Wiley.

Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research, 20,* 241–254.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1.*

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tori, C. D. (1989). Quality assurance standards in the education of psychologists: Reliability and validity of objective comprehensive examinations developed at a freestanding professional school. *Professional Psychology: Research and Practice, 20,* 203–208.

Vernon, P. E. (1961). *The structure of human abilities.* Andover Hants, England: International Thompson Publishing Services.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8,* 216–226.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81,* 557–574.

Watson, G. B., & Glaser, E. M. (1980). *WGCTA®/Watson Glaser Critical Thinking Appraisal Forms A and B.* San Antonio, TX: The Psychological Corporation.

Watters, G. V., & Paterson, D. G. (1953). Miller Analogies Test scores and ratings of PhDs in psychology. *American Psychologist, 8,* 89–90.

Wechsler, D. (1946). *The Wechsler–Bellevue Intelligence Scale, Form II. Manual for administering and scoring the test.* New York: The Psychological Corporation.

Yerkes, R. M. (Ed.). (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences, 15,* 1–890.

Appendix

Corrections to Criterion-Related Validities

Range-Restriction Corrections

To correct for range restriction, the ratios of selected group standard deviations to applicant-pool standard deviations (*u* values) are necessary. Graduate student applicant standard deviations were available from the publisher of the MAT (The Psychological Corporation, 2001). Standard deviations were available for eight program areas as well as an overall standard deviation. Because MAT-applicant standard deviations differed by program area, they were matched by area with the sample standard deviations. This was done to avoid over- or undercorrecting the observed correlations and to more accurately represent the true variability of range-restricted samples in the literature. Two range-restriction distributions were created. One comprised MAT standard deviations from student samples with criteria from earlier stages of graduate programs (1st-year GPA, GGPA, faculty ratings) , whereas the second was based on student samples with criteria from later in graduate programs (degree attainment, research productivity). This was done to better reflect the effects of attrition on graduate student samples and better match the correction to the sample (although the difference between the corrections was small). Standard deviations were not available for either job applicant or incumbent samples. Because most of the studies that contributed to the job performance criteria were for jobs with moderate to high levels of complexity and many had workers that would have attended college (e.g., counselors, managers, engineers), it was assumed that the restriction of range-artifact distribution for the student samples would result in a reasonable correction for the studies in work settings. However, this correction may be an undercorrection because our *u* values are smaller than those typically obtained in the work and even graduate school domains (Kuncel et al., 2001).

Criterion Reliability Corrections

The reliability of grades was based on reliabilities from three studies of the reliability of college grades from Reilly and Warech (1993), Barritt (1966), and Bendig (1953). The values for internal consistency reliability from these three studies were .84, .84, and .80, respectively. Note that unlike typical internal consistency estimates that are based on a single test administered at one point in time, estimates are from final grades (typically based on multiple subevaluations) from multiple raters over the course of months or years.

The artifact distribution for the reliability of faculty ratings was estimated in two stages. First, available reliability information about the number of raters and number of rating items was coded from the MAT articles. Most studies provided information about the number of faculty ratings and items. Unfortunately, almost no studies provided an estimate of the interrater reliability. To compensate for the lack of information, a meta-analytically estimated reliability for a single rater rating overall job performance was obtained from Viswesvaran, Ones, and Schmidt (1996). This estimate (.52) was then used to create the artifact distribution after being adjusted for the number of raters. Because Viswesvaran et al.'s estimate was for a single rater across different numbers of items and our information suggested that there was also a range of number of items in our sample, no adjustment was made for the number of items. To the extent that there are differences between the average number of items in Viswesvaran et al.'s meta-analysis and our study (average number of items = 2), our correction may lead to somewhat of an over- or underestimate.

The same method was used to estimate the values in the artifact distributions for the reliability of potential and creativity ratings, with one exception. One large sample study of creativity provided an interrater reliability estimate. This estimate was incorporated in the analysis without adjustments. Because of the overlap between the professional potential criterion analysis and the counseling potential analysis, the potential rating distribution was also used for the counseling potential criterion. Note that all of the potential ratings for which we have information were made with a single item. Therefore, the correction is likely to be a slight undercorrection because the meta-analytic estimate used here was based on a single rater across a range of items.

In the studies we obtained, there was almost no information about the reliability of the job performance, counseling performance, and educational administration criterion measures. The reliability distribution for these criteria was based on the meta-analytically derived reliability estimate for a single overall job performance rating (.52) from Viswesvaran et al. (1996) adjusted for the number of raters with one exception. The exception is for one job performance study where the criterion was an objective measure that was based on records. We conservatively assumed the criterion reliability for this study was 1.0 and included it in the distribution. All others were estimated based on the number of raters when that information was available. Across all ratings-based criteria, there were typically a large number of raters, hence the comparatively large average reliability of .81.

The reliability distribution for work samples was constructed from 18 individual studies reporting internal consistency (mostly split half) reliabilities from the industrial and organizational psychology literature. Work sample performance is frequently used as a criterion variable in validating selection tests or as a proximal predictor of future performance. Previously, in modeling determinants of performance, Schmidt et al. (1986) constructed a reliability distribution for work samples from the literature available at the time. We included all reliabilities used by Schmidt et al. (1986) but also updated their distribution using data from studies conducted since then. The average reliability of work samples was estimated to be .88, a value almost identical to the mean value of .87 reported by Schmidt et al. (1986).

No reliability corrections were made for comprehensive examination scores, degree attainment, time to finish, research productivity, number of course credits, and membership in a professional organization. The criteria that were obtained from records or databases (membership, graduation,

credits, publications) are likely to be highly reliable, and the meta-analytic results presented here are likely to closely reflect the true validity of the MAT for these criteria.

Comprehensive examination scores are much less likely to be highly reliable. Although some adequate split-half reliabilities have been reported for multiple-choice comprehensive examinations (Tori, 1989), scoring of written comprehensive examinations may often rely on the idiosyncratic judgments of raters. The format of comprehensive exams (essay vs. multiple choice), the number of raters for scoring essays, and the amount of rater training can all be expected to moderate the reliabilities of comprehensive examination scores. Unfortunately, detailed information regarding the nature of comprehensive examinations was not provided in all of the studies reporting correlations with the MAT. Therefore, we felt we could not develop an artifact distribution that appropriately reflected the distribution of reliabilities. We expect that the results presented here are likely to grossly underestimate the validity of the MAT for predicting comprehensive examinations.

MAT Reliability Corrections

Test–retest alternate-form reliabilities were collected from the technical manuals and the literature to create a distribution of predictor reliabilities. To help reduce the possible effects of maturation on the reliability estimates, readministration of the MAT had to occur during a time period of less than a year to be included in this artifact distribution. Artifact distribution information for range restriction, predictor reliability, and criterion reliability corrections are presented in Table A1.

Corrections for Correlations With Cognitive Tests

Unfortunately, most of the studies reporting correlations between the MAT and other cognitive ability tests did not provide sufficient information to make test-specific restriction of range corrections. Therefore, we corrected for the restriction of range using the information we compiled for the MAT. Test manual reliability information was available for many of the other tests and three separate reliability distributions were created for the GRE–V, GRE–Q, and other non-MAT tests. To summarize, the analyses were conducted with the MAT reliability distribution, a separate non-MAT reliability distribution, and restriction of range corrections that were based on the MAT range-restriction artifact distribution for the graduate student samples.

Table A1
Statistical Artifact Distributions Used in the Meta-Analyses

Variable	Mean U_{RR}	K_{RR}	Mean $r_{XX}^{1/2}$	K_{rel}
Predictor				
Miller Analogies Test[a]	.83	26	.92	6
Miller Analogies Test[b]	.76	8		
Criteria				
GGPA	—	—	.91	3
1st-year GGPA	—	—	.91	3
Faculty ratings	—	—	.86	18
Counseling work sample	—	—	.88	18
Creativity ratings	—	—	.87	5
Potential ratings	—	—	.83	6
Job performance	—	—	.81	6

Note. Mean U_{RR} = mean U ratio for range restriction; K_{RR} = number of ratios in the distribution; Mean $r_{XX}^{1/2}$ = mean of square root of the reliabilities (multiple raters for ratings criteria); K_{rel} = number of reliabilities in the distribution; GGPA = graduate grade point average.
[a] Less range-restricted artifact distribution based on students early in their programs. [b] More range-restricted artifact distribution based on students late in their programs.

Received October 1, 2002
Revision received April 17, 2003
Accepted May 6, 2003 ■