
A Meta-Analysis of the Predictive Validity of the Graduate Management Admission Test (GMAT) and Undergraduate Grade Point Average (UGPA) for Graduate Student Academic Performance

NATHAN R. KUNCEL
University of Minnesota

MARCUS CREDE
LISA L. THOMAS
University of Illinois at Urbana-Champaign

Considerable debate both within and outside of academic circles surrounds the validity of standardized tests for predicting student performance in graduate business schools. This meta-analysis aggregates the existing literature on the validity of the two most heavily used predictors in business school admissions: the GMAT and undergraduate grade point average. Results based on over 402 independent samples across 64,583 students indicate that the GMAT is a superior predictor to UGPA and that the two combined yield a high level of validity for predicting student performance.

One of the most widely used predictors of business student performance is the Graduate Management Admission Test (GMAT). Recent surveys have found that approximately 1,700 schools use the GMAT for admissions and more than 1,000 require the GMAT for admissions. Over the past 5 years, an average of over 190,000 tests were administered each year (GMAC, 2003), and considerable re-

search has examined the predictive validity of the GMAT for predicting graduate student performance in business programs. Unfortunately, there is also considerable disagreement about the effectiveness of the GMAT. Previous researchers have variously argued in favor of the GMAT (e.g., Bieker, 1996; Melnick, 1972; Morris, 1995) or against it (e.g., Grambsch, 1981; Goodrich, 1975; Hancock, 1999; Fairtest, 2003; and see Fisher & Resnick, 1990 for a review). This is part of the larger on-going controversy over the use of standardized testing for admissions (Zwick, 2002).

We would like to thank David Klieger, Stephanie Seiler, and Sang Eun Woo for their proofing the data coding and their helpful comments and suggestions. We would like to thank Meghan Dalton, Edward Hezlett, Elizabeth Varones, and Patrick Walsh for their assistance in gathering the articles summarized in this meta-analysis and Lisa Maillie and Jennie Cordis for their assistance with manuscript preparation.

Correspondence concerning this article should be addressed to Nathan R. Kuncel, Department of Psychology, University of Minnesota, Minneapolis, MN 55455. E-mail: kuncel001@umn.edu

Some of the controversy is fueled at least in part by the range of predictive validities observed in the literature from a low of $-.45$ to a high of $.76$ (Dobson, Krapljan-Barr, & Vielba, 1999; Bottger & Yetton, 1982). Predictive validity is certainly a critical concern, especially when considered relative to other options. There is little to discuss if a pre-

dictor simply does not predict. However, the issues surrounding the GMAT go beyond validity because of the complexity of admissions systems and their components. Admissions systems can be broken into two major parts. The first are the predictors or measures used to forecast future student performance. The second is the method by which the predictors are combined to make admission decisions. These two parts result in concerns about the validity of admissions predictors, their relationships with other predictors, and their use in applied settings. In addition, the consequences of an admissions system are also important, and this raises the issue of predictor bias and fairness for each component and method used to combine the information for decision making. At stake are the quality and the diversity of MBA students.

Our primary purpose here is to address the first issue, predictive validity, by quantitatively aggregating the available literature on the two most consistently used predictors: the GMAT and undergraduate GPA. Our study combines the available information on the GMAT into a single study and provides a single comprehensive piece of validity evidence. With validity evidence in hand, we then discuss the remaining issues associated with the development of admissions systems and the use of standardized tests. To facilitate interpretation of the validity results and the discussion of admission system design, we first present a conceptual framework.

Framing Admissions Measures

It is useful to frame the content of admissions predictors in terms of the classic distinction between *signs* and *samples* made by Wernimont and Campbell (1968) in their paper titled, "Signs, Samples, and Criteria." The fundamental goal of selection or admissions is to predict who will engage in a set of important behaviors either on the job or at school. These important behaviors are variously labeled performance or achievement, but regardless of what we call them, they are the criteria of interest. The measures we use to predict these important criteria can be roughly categorized as either signs or samples. A *sample* is a direct measure of the actual criterion behaviors of interest. In contrast, a *sign* is a measure that tends to be associated with the actual behaviors of interest without directly measuring them.

The current debate over the use of prior work experience in MBA admissions is a perfect practical illustration of the difference between signs versus samples. A number of MBA programs require or consider prior work experience because they

believe that it will lead to positive student behaviors including students who can better understand the business setting and more readily assimilate their education (Dreher & Ryan, 2000; Dreher & Ryan, 2004). However, experience is purely an indicator (*sign*) and not a direct measurement (*sample*) of the actual knowledge, skill, abilities, and other characteristics (KSAs) the program wishes the students to possess. That is, using prior experience is measuring something that is only associated with the actual characteristics desired by the program and not directly quantifying the desired characteristics.

Ideally the focus of admissions should be on direct measurement of the desired knowledge, skill, abilities, and other characteristics and not on signs. All else being equal, a sample should be maximally valid, fair, and unbiased because it quantifies the actual behaviors of interest. Finally, a sample is likely to be perceived as face valid and be well received by applicants because it, again, directly reflects the KSAs of interest. Note that use of samples for admissions have the additional advantage of forcing organizations to specify what they believe constitutes effective learning and behavior, a healthy exercise for any MBA program.

Signs are particularly undesirable under two conditions. Signs are less desirable predictors when it is possible to acquire the desired characteristics without fulfilling the sign. For example, students may very well have the seasoning and insight associated with work experience through other nonwork experiences or training. Signs are also less desirable predictors when the connection between the sign and the desired characteristics or behaviors is tenuous. For example, most MBA programs have not clearly and carefully specified what it is that a year of work is supposed to yield. Even with such a specification, time spent on the job is a very poor substitute for direct measurement of the relevant characteristics. Not all work experiences are equal in quality, and it is certainly possible for one person to have 5 years of experience while another has 1 year of experience 5 times.

This is not to say that signs are always inappropriate. Signs are generally more acceptable under three conditions. The first is when they are particularly robust predictors. That is, a sign that works well, is low in cost, and contributes to the overall validity of the admission system is, all else being equal, probably worth using.

The second is when the costs associated with obtaining a more direct and high-fidelity sample are so high its utility is diminished. An extreme example can illustrate this point. The best possible

predictor of MBA student success we can dream up would be to set up a simulated business school with all of its associated facilities (e.g., libraries, computer labs) and distractions (e.g., bars, social opportunities) and then have students attend it for a semester. This hypothetical "MBA Biosphere" would not, of course, be a feasible admission tool because the costs would equal those of actual MBA education. However, it would be an exceptional predictor of both early and later MBA performance and a clear example of a sample. Note that development of this sample would again have the additional desirable feature of forcing educators to specify what criteria constitute effective performance.

Finally, the third condition under which signs are more appropriate is when a great deal of learning will occur post hire or admission. Any sign that is strongly predictive of the ability to learn would, all else being equal, be a useful addition to an admissions system. This point is especially salient for the current educational context.

In work settings we often select people who may not fully know how to do the job and display all of the needed behaviors for the job starting their first day at work. School settings are a more extreme case in that it is assumed that students are not fully knowledgeable and skilled. Teaching the necessary knowledge and skill is, after all, the purpose of education. In this case, use of samples for predicting subsequent success may be difficult and inappropriate.

Instead, we will need to sample learning skills or amount of prior learning rather than the specific behaviors students are expected to have mastered after they have finished various stages of their education. This can take the form of predictors of learning that range from domain specific to domain general. A domain-specific measure could be a measure of necessary prior knowledge or interest in a specific topic. A domain-general measure would broadly sample prior learning or motivation to learn in general. Samples of prior learning can be effective because they tend to be predictive of subsequent learning.

Examining more traditional predictors within this framework helps clarify the likely sources of their validity. For example, students' undergraduate records can be viewed as imperfect samples. Undergraduate training is not identical to graduate training, and the behaviors exhibited that lead to effective performance in college are not the same as those needed for effective graduate performance. However, there is considerable overlap, and prior grades are consistently good predictors

of performance. In addition, prior grades have increased utility because they are quite low in cost.

We view the GMAT as both a sample and a sign. The GMAT is a sample in that it quantifies a wide range of specific KSAs that are very important for student performance. The ability to read and extract information, fundamental math skills, analytical writing, and data interpretation are all necessary for successful performance in MBA programs. The GMAT is also a sign because it measures a wide range of prior learning, much of which is not highly domain specific. It is not highly domain specific because the GMAT, quite appropriately, does not touch on knowledge that is likely to be taught in an MBA program. The GMAT, like all ability measures, is an important sign of the ability to learn and acquire skill.

Empirical and Theoretical Background

We can conceptualize general cognitive ability as the "repertoire of intellectual (or cognitive) skills available to the person at a particular point in time" (Humphreys, 1989: 194) and define cognitive ability measures or tests as being samples of this repertoire. Using this definition from Humphreys, (1989) we can see the connection between cognitive ability measures and subsequent learning and performance more clearly. Individuals who have acquired a larger and more advanced repertoire are more likely to already have the needed knowledge and skill necessary for academic tasks. The behaviors measured by the GMAT share key elements with the behaviors students need to exhibit to be successful. Furthermore, students are also more likely, on average, to be able to acquire additional knowledge and skill in a specific domain because prior domain-specific knowledge generally facilitates additional acquisition. For example, a considerable command of geometry can facilitate learning more complex math and statistics.

The following sections provide empirical evidence supporting the relationship between cognitive ability and academic success, and training success. Finally, we summarize with empirically supported models of the relations among ability, knowledge, skill, and performance.

Cognitive Ability and Graduate Student Performance

The first, and most directly relevant, evidence supporting the validity of the GMAT are those studies that have examined the relationship between measures of cognitive ability and academic performance. In large-scale validity studies or meta-

analyses, cognitive ability measures such as the LSAT, GRE, MAT, PCAT, and MCAT have been found to be valid predictors of multiple aspects of student success in law school (Linn & Hastings, 1984), masters and doctoral graduate studies (Kuncel, Hezlett, & Ones, 2001, 2004), pharmacy programs (Kuncel, Crede, Thomas, Seiler, Klieger, & Woo, 2005), and medical schools (Mitchell, 1990), respectively. Based on the validity evidence and the interconnection among cognitive ability measures, it is reasonable to expect that the structurally similar GMAT would be predictive of academic success in business school.

Cognitive Ability and Formal Training Programs

The second, and more generally relevant, evidence is the considerable body of research that establishes a link between cognitive ability measures and learning in general. Research has found substantial positive relationships between measures of cognitive ability and skill acquisition (e.g., Lohman, 1999; Ackerman, 1987, 1992). Similarly, cognitive ability has been found to predict training success in employment settings (Ree, Caretta, & Earles, 1998; Schmidt, 2002). Both educational and occupational domains require the acquisition of new knowledge and skill. In formal training programs and degree programs, the courses are structured interventions designed to increase the knowledge, skill, or motivational determinants of performance (Campbell & Kuncel, 2001). The need to acquire knowledge and skill both formally and informally in work and educational settings as well as apply previously acquired knowledge and skill to perform tasks provides one major explanation for why cognitive ability measures predict performance in both educational and work settings. The relations among ability, skills, knowledge, and performance have been integrated into job performance models that we believe are useful for understanding the GMAT.

Models of Individual Differences and Job Performance Relationships

A series of models have specified how individual differences (e.g., cognitive ability, conscientiousness) are related to performance. Work by Campbell and colleagues has organized individual differences into direct and indirect determinants of performance (Campbell, Gasser, & Oswald, 1996; McCloy, Campbell, & Cudeck, 1994). The direct determinants of each dimension of performance are limited to procedural knowledge (i.e., job knowledge), declarative knowledge (i.e., skill), and voli-

tional choices (i.e., motivation). Campbell and colleagues argue that the direct determinants are the most proximal determinants of performance. Furthermore, other individual differences such as general cognitive ability, experience, and personality traits are said to influence performance through their influence on the direct determinants. For cognitive ability, the mechanism for the indirect influence is through the acquisition of job-specific knowledge and skill. For example, having higher levels of cognitive ability would be associated with having higher levels of job knowledge in the area of finance and skill analyzing financial data, which, in turn, would directly result in superior job performance in developing models to determine stock investing for a mutual fund. In other words, job knowledge and skill nearly fully mediate the relationship between general cognitive ability and job performance. The major conclusion is that general cognitive ability influences the acquisition of knowledge and skill which, when combined with motivation, are directly related to actual performance.

Similar findings have been obtained in the academic domain. Kuncel et al. (2001) found that the specific GRE subject tests were consistently better predictors of seven different measures of graduate student performance than prior grades and the GRE-Verbal and GRE-Quantitative. The more general tests were also valid predictors of performance but to a lesser extent than the subject tests, even though subject tests and general tests are strongly correlated. Therefore the more specific "job knowledge" measure in the form of the subject test nearly fully mediated the relationship between the general tests and academic performance. Based on this body of research, we would also expect that GMAT items measure both direct and indirect determinants of performance (which can be likened to samples and signs). First, the GMAT measures abilities related to the acquisition of relevant knowledge and skill (as discussed above). Second, the GMAT directly samples knowledge and skill that will be of immediate use and importance in business school (e.g., mathematical reasoning skill that is relevant for interpreting financial business data, basic writing and English skills that are needed for writing papers and reports).

Cognitive Ability and Managerial Performance

The nature the student's job and its relationship to the work obtained by many recipients of MBAs is important because if the GMAT predicts student success it should also have a relationship with

subsequent job success. That is, if there are task similarities between the two domains, individual differences that predict in one context are likely to be valid in others. Research on managerial performance has demonstrated the validity of cognitive ability measures for predicting performance in managerial jobs.

In their classic review of the managerial performance literature, Campbell, Dunnette, Lawler, & Weick (1970: 129) concluded, "It is apparent that proficiency in executive and managerial jobs has been most effectively indicated by tests of intellectual ability, perceptual accuracy, and personality or interest." Similarly, in his review and synthesis of the managerial selection research Cascio (1998: 226) commented, "The fact is, general cognitive ability is a powerful predictor of job performance." Measures of cognitive ability are related to the acquisition of job knowledge just as they are related to the acquisition of knowledge needed in MBA programs. Some may draw a large distinction between academic work and the work of managers. We would like to suggest that the task content of both jobs is less dissimilar than some may believe.

Substantive Nature of Business School Performance

We can conceptualize effective performance in a graduate business school program as, fundamentally, the demonstration of recently acquired knowledge and skill. Students come to business school to acquire knowledge and skills that are relevant for performance in business positions including finance, IT, marketing, and human resource management. During their time in business school, students engage in a wide range of behaviors that are directly and indirectly relevant for acquiring this knowledge and skill. Successful performance is typically measured through an evaluation of one or more final products including examinations, papers, presentations, and in-class case studies.

However, when considering student performance, it is important to remember that in-class performance on papers, presentations, and examinations are the *end product* of a great deal of prior performance. In other words, when thinking about potential predictors of graduate student performance, we have to realize that students engage in a great deal of complex behavior both within and outside of the classroom. Ultimately, this behavior has an influence on the quality of the directly evaluated traditional classroom behaviors (e.g., examinations, presentations, term papers). Most of the

behavior that indirectly influences the quality of the end products goes unobserved by the faculty.

A number of studies have been designed to examine the structure of graduate student performance, and they all indicate complex models that include many domains of complex behavior that are not consistent with the notion that academic performance is substantially different from job performance (Enright & Gitomer, 1989; Reilly, 1976; Walpole, Burton, Kanyi, & Jackenthal, 2001; and see also Kuncel, 2003 for a review). The nature of the performance of business school students is likely to share many dimensions with these more general studies. Across all studies, graduate student performance is found to be complex and multidimensional. Many of its dimensions (e.g., time management, resolving goal conflicts, extracurricular leadership) occur outside of the classroom and involve problem-solving tasks not unlike those found on jobs. This more complex conceptualization of the performance of students in business school has important implications for the prediction of performance. A predictor can be valid, because it directly correlates with the behaviors evaluated in the end products (e.g., answers to examination items), or a valid predictor may be correlated with effective behaviors that, in turn, lead to the observed end-product behaviors. It is necessary to consider what characteristics appear to be relevant for performance on the end product as well as the individual difference characteristics that may be relevant for all of the behaviors that lead up to (or support) effective performance on the end product.

Numerous specific behaviors will have both direct and indirect influences on traditional classroom success. These include time management, negotiating with peers, seeking additional information for study, self-assessments of current levels of knowledge in an area, avoiding counterproductive behavior (e.g., alcohol use, procrastination), and even dealing with other aspects of student life including managing school- and life-related finances. Students do many things during the semester that facilitate their arrival at the final examination with a mind filled with new knowledge; therefore, any performance on an examination is the function of a great deal of prior complex behavior.

The line of reasoning we present here is at odds with previous conceptual discussions of academic versus practical tasks (e.g., Sternberg & Wagner, 1993). Academic tasks are said to be well defined, require collection of no additional information, and have one correct answer, among other things. Although a few individual tasks (like taking a multiple choice test) may fit the definition of an aca-

ademic task, previous discussions ignore the complex behaviors that lead up to the test. That is, they do not begin to consider the behaviors that, ultimately, result in individual differences in performance on the classroom end product.

We feel that this more complex view of academic performance is particularly true for graduate students. In business schools, MBA students engage in many complex tasks both within and outside of the classroom. Business programs often attempt to simulate complex and ill-defined business problems in a group setting that requires information gathering, teamwork, and complex problem solving for ill-defined problems (i.e., the business case study). Business school performance has many parallels with job performance, and we should expect findings and models for job performance to have relevance for understanding academic performance. Indeed, previous research has found that an academic admissions test is a valid predictor of actual job performance and creativity (Kuncel et al., 2004). That is, research has demonstrated that ability measures are related to performance in both domains. Consistent with this body of research, we expect that the GMAT will predict evaluations of the complex behaviors that occur in business school.

Criteria Examined and Hypotheses

Since the scales in the GMAT measure broadly important skills and knowledge that have been acquired (i.e., learned) over an extended period of time, we would expect them to be related to subsequent performance and learning as represented by first-year grade point average and overall graduate grade point average. As noted above, these general abilities are related to acquiring job knowledge, skill acquisition, and job performance. Performance in both settings is predicted by measures of cognitive ability.

Hypothesis 1a: The GMAT will be a valid predictor of grade measures of student performance in business school.

Hypothesis 1b: Undergraduate grade point average will be a valid predictor of grade measures of student performance in business school.

We also obtained a number of studies that examined the predictive validity of grades earned in the junior and senior years of college. Overall, we might expect that grades obtained closer in time to graduate school would, all else being equal, be somewhat better predictors. This hypothesis is consistent with research that has found that, on average, the closer in time grades are obtained the

larger the correlation between grades (Humphreys & Taber, 1973; Humphreys, 1968).

Hypothesis 2: Junior/senior college grade point averages will have slightly better validity than overall undergraduate records.

Although we might expect that prior performance in the form of undergraduate grade point averages will be the best predictor of subsequent graduate student performance, this has not been found to be the case. Both Kuncel et al. (2001), and Linn and Hastings (1984) found that undergraduate records are generally not better predictors than standardized tests for predicting various measures of performance in graduate school. This may be due to grade inflation effects or to the fact that undergraduate grades from liberal arts institutions are based on a range of topics, some of which are very different from the specific topics studied in graduate school. Therefore, we expect undergraduate grade point average to be valid but not to a greater extent than the GMAT.

Hypothesis 3: The GMAT will have equal or greater validity than undergraduate record for predicting graduate student grades and GPAs.

The demographic composition of GMAT test takers has changed substantially over time. Recently, the number of non-U.S. students taking the GMAT has increased (GMAC, 2003). Many of these test takers are nonnative English speakers. Although there is evidence that standardized tests remain valid predictors for nonnative English speakers (Kuncel et al., 2001), this demographic change remains a concern for the GMAT. Based on previous research, we expect the GMAT to be valid for predicting the performance of nonnative English-speaking students. In particular, we expect the less verbally saturated quantitative scale to be more predictive than the verbal scale of performance in business school for nonnative English speakers. Students may arrive at school with relatively poor verbal skills but, due to immersion in a new setting, may rapidly acquire additional skills. Furthermore, nonnative students may emphasize less verbally loaded disciplines (economics, finance) than would native English speakers.

Hypothesis 4: The GMAT will be a valid predictor of grade performance in business school for nonnative English-speaking students with superior validity for the Quantitative scale.

Persisting in graduate school is an important outcome. Students who stay in the program and obtain the degree will generally have more career options. Generally, students who are admitted into

a program have the capability to finish it. Therefore, we would expect persistence to be largely determined by drive and motivation. Finishing a degree is heavily a product of consistently choosing to study, studying intently, and persisting at one's studies and homework over time. However, ability still does play some roles, as previous researchers have found modest but positive relationships between ability measures and persistence outcomes in graduate school (Kuncel et al., 2001, 2004). Therefore, we expect the GMAT to have a modest but positive relationship with persistence in business school.

Hypothesis 5: The GMAT will have a modest but positive correlation with persistence in graduate school.

It is possible that other variables moderate the validity of the GMAT. Important to note is that the moderation may be the result of the *criterion* (i.e., course grades) rather than of the predictor. For example, at the undergraduate level, the genders have disproportionately historically pursued different majors that have different stringent grading standards (Young, 1991). In this study, sufficient information was available to examine 5 additional moderators that are of applied importance.

The first was validity for male versus female students. This is, of course, a large concern. Based on previous research, we would expect the validity to be comparable for men and women (McKelpin, 1965; Stanely & Porter, 1967). Undergraduate educational background was examined with two different moderators. Specifically, we contrasted the validity for undergraduate business major versus nonbusiness majors and studies that examined quantitative undergraduate degree holders versus those with nonquantitative degrees. We might expect lower predictive validities for those students with a business background because the MBA program places less of a demand on their ability to acquire new knowledge. However, many other undergraduate degrees will also provide important background knowledge. In addition, this picture of validity may be partially obscured by differential restriction of range on the predictor due to ability differences across undergraduate majors (Shuey, 1950; ETS, 2003). Therefore, we would expect the validities to be very similar across these different backgrounds with, perhaps, slightly lower correlations for students with undergraduate business degrees. Finally, we have sufficient data to examine the validity of the GMAT for full- versus part-time students. We expect the effects of these moderators will be minor because the nature of the work to be done in both cases is very similar. As such, the GMAT should remain comparably valid. However,

grading differences across program types and sample differences may produce some modest validity difference.

Hypothesis 6: The validity of the GMAT will not be meaningfully moderated by gender, full- or part-time students, or the type of undergraduate degree held by students.

Across all analyses, data from 402 independent samples yielded a total of 1,639 correlations across 64,583 students. Each analysis consists of a unique set of individuals, and independence was not violated in any analysis.

METHODS

The data collected from the studies were analyzed using the Hunter and Schmidt (1990) psychometric meta-analytic method. This method was preferred above others, because it permits estimation of the amount of variance attributable to sampling error, range restriction, and unreliability. We used the artifact distributions described later to correct for the attenuating influences of artifacts on the observed correlations. The interactive meta-analysis procedure was used (Hunter & Schmidt, 1990: 165; Schmidt, Gast-Rosenberg, & Hunter, 1980) to improve the accuracy of the results.

A meta-analysis is particularly desirable for several reasons. First, when evaluating a literature on the validity of a predictor, it is important to remember that the results from individual studies are affected by sampling error and may, by chance, vary widely from the actual validity. The effects of sampling error are more likely to be strong when studies are conducted with small sample sizes (Hunter & Schmitt, 1990). A common finding is that much of the variability across studies can be accounted for by sampling error. Observing some studies with poor or even negative validities is to be expected, given that many of the studies done on the GMAT have very small sample sizes. For example, the average sample size of the studies contributing to the validity estimates for GMAT-Verbal and first-year GPA is 144, which, when examined in isolation, can lead to misleading conclusions about the GMAT. When studies are aggregated, the total sample size is much larger and results in a far more accurate evaluation of the predictive validity.

Description of the Database

We gathered studies involving the GMAT from several sources. To identify relevant research, PsychLIT (1887–2003), EconLIT (1969–2003), Health and

Psychosocial Instruments (1985–2004), ABI/Inform Complete (1971–2004), Education Full Text (1994–2004), and ERIC (1966–2003) searches were combined with a search of Dissertation Abstracts International (1861–2002). The authors discussed search terms, and all three authors conducted searches. Very broad searches were conducted and were subsequently reduced by inspecting each abstract. Inspecting abstracts suggested some new key word searches and these were added to the original searches. After the initial set of promising sources were identified and collected, the citation lists within these articles, dissertations, and technical reports were also examined to identify additional relevant studies. Unreported effect sizes were computed from available information when possible. Finally, hand searches were conducted of the *Journal of Education for Business* and *Educational and Psychological Measurement* to identify studies missed by search terms. In total, the data presented in this study was obtained from 9 dissertations, 25 journal articles, and 12 technical reports.

To address potential overlaps between data reported in different sources, text searches by author name were conducted. Studies with identical authors were examined. In articles with sample overlaps, the larger or more complete data were included in the meta-analysis, and the matching articles were excluded.

Intercorrelations among the predictor variables were also coded. These data were obtained to permit estimation of incremental validity. Some data obtained were for preadmission samples, while others were for incumbent samples. These data were aggregated after the preadmission correlations were corrected for restriction of range in both variables.

Measures

The GMAT, written and administered by the Educational Testing Service (ETS), assesses an individual's qualifications for advanced business and management training. Administered since 1954 via a paper-and-pencil format, the majority of test-takers now take the Computer-Adaptive GMAT. The test consists of the following four timed sections: Quantitative, Verbal, and two Analytical Writing sections. The Analytical Writing section contains the following two separately timed parts: Analysis of an Issue and Analysis of an Argument. The Quantitative section consists of items labeled "Data Sufficiency" and "Problem Solving." The Verbal section contains "Reading Comprehension," "Critical Reasoning," and "Sentence Correction" scales. The primary purpose of the GMAT is to predict grades.

Criterion Measures

Studies examined in the analyses of first-year graduate grade point average all contained first-year graduate business school grades except for one study that presented GPAs in foundation and functional core courses. These were averaged together and included in the first-year GPA analyses.

For the overall graduate grade point average criterion, most studies were comprised of complete graduate GPAs. The only two exceptions included one study that specified second-year GPA but was ambiguous about whether the GPA was after the second year or just during the second year (Powers & Moss, 1980). Given that the goal of these analyses is to examine grades after the first year, this study was included in the analyses. A second study included in the meta-analyses examined GPA after at least one class was completed by a student (Schwan, 1988). In other words, students who dropped out were in this study's analyses. Given that most students do persist through business programs, this study was also included as it provided some information about grade performance after the first year.

Persistence and drive were defined somewhat differently across all studies but still reflect making progress through the graduate program. Persistence was defined as graduating versus failing to graduate from the program, staying versus dropping out after the first year, completing a thesis versus failing to do so, and completion of the MBA versus failure to complete it. All variables were dichotomous. Overall, these studies are related to the general outcomes of persistence and degree attainment and were combined to provide information about this important outcome.

Range Restriction Distribution

Selection on the basis of a variable reduces the variance in the predictor and results in attenuated correlations with criteria (assuming the relationship is not zero; Thorndike, 1949). The effects of range restriction can be addressed through appropriate corrections. Defining the population of interest is critical when correcting for range restriction. In our study, all potential applicants to a graduate program or job were considered to be the population of interest. Therefore, the population estimate was taken as the population of test-takers. To correct for range restriction, the ratios of selected group standard deviations to applicant pool standard deviations (i.e., U values) are necessary. Graduate student test-taker standard deviations

were available from the Graduate Management Admissions Council through their annual publication of the *Guide to the Use of GMAT Scores*.

Since GMAT applicant standard deviations differed slightly over time, they were matched with the sample by year. 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. Separate distributions were created for each scale. Since no population level standard deviations are available for undergraduate grade point average, the combined distribution for GMAT-Verbal and GMAT-Quantitative was substituted to approximate a reasonable restriction correction. Of course, when multiple predictors are involved, it is a multivariate restriction of range issue (Aitken, 1934; Lawley, 1943; Sackett & Yang, 2000). Unfortunately, conducting such corrections is not possible with most meta-analytic data. Furthermore, the predictors are strongly intercorrelated, and we have no reason to believe that dramatically different weighting schemes are used in admissions. Therefore, we expect that the current corrections will result in more accurate results than would no corrections at all. Uncorrected sample size weighted average correlations are also presented and demonstrate the same pattern of relationships.

We also obtained a particularly strong set of range restriction ratios (i.e., U values) for GMAT-Total scores. It is unclear if these are due to the particular samples for which we had sample standard deviation estimates or if they may be some artifact of the admissions decision process. These corrections result in validity estimates that are larger than what we would expect from the unit weighted combination of the GMAT-Verbal and GMAT-Quantitative validities.

Criterion Reliability Distributions

Since the correlation of interest is between the tests and graduate school performance, the reliability of the measure of graduate school performance is an issue. The unreliability of the grade measures lowers the observed correlation between the performance measure and the GMAT. The reliability of grades was based on internal consistency reliabilities from three studies of college grades from Reilly and Warech (1993), Barritt (1966), and Bendig (1953). Artifact distribution information for range restriction, predictor reliability, and criterion reliability corrections are presented in Table 1.

Note that unlike typical internal consistency estimates based on a single test administered at one

TABLE 1
Artifact Distributions Used in the Meta-Analyses

	MU_{RR}	k_{RR}	$M r_{xx}^{1/2}$	k_{rel}
Predictors				
GMAT - Verbal	.73	264		
GMAT - Quantitative	.83	263		
GMAT - Total	.70	61		
UGPA	.78	527		
Criteria				
GGPA	—	—	.91	3
first-year GGPA	—	—	.91	3

Note. $M U_{RR}$ = mean U ratio for range restriction; k_{RR} = number of ratios in the distribution; $M r_{xx}^{1/2}$ = mean of square root of the reliabilities (multiple raters for ratings criteria); k_{rel} = number of reliabilities in the distribution.

point in time, estimates are from final grades (typically based on multiple subevaluations) from multiple raters over the course of months or years. Since persistence criteria come from records, they were assumed to be perfectly reliable, and no criterion reliability corrections were made.

RESULTS

Results for overall analyses are presented in Table 2. The GMAT and UGPA were hypothesized to be valid predictors. The results support these hypotheses. The validities for First-Year Graduate Grade Point Average were moderately large for the GMAT-Verbal ($N = 48,915$, $k = 315$), GMAT-Quantitative ($N = 48,758$, $k = 314$), GMAT-Total ($N = 28,624$, $k = 202$), and UGPA ($N = 50,138$, $k = 324$) with operational validities of .34, .38, .47, and .35, respectively. The standard deviation of these true validities ($SD_{\rho\rho}$) was small compared to previous meta-analyses on standardized tests. We also hypothesized that the GMAT would have equal or greater validity than the UGPA. Results support this, with the GMAT total exceeding the validity for UGPA and the individual scales averaging .0125 less than the UGPA.

For Graduate Grade Point Average, the validities were moderately large for the GMAT-Verbal ($N = 5,466$, $k = 28$), GMAT-Quantitative ($N = 5,609$, $k = 27$), GMAT-Total ($N = 5,201$, $k = 29$), and UGPA ($N = 5,609$, $k = 28$) with operational validities of .32, .30, .47, and .35, respectively. We hypothesized that Junior/Senior college grade point average would be a stronger predictor than overall UGPA. This result was not supported by the results, with a validity of .31 ($N = 1,292$, $k = 7$) for Junior/Senior GPA.

The hypothesis for the persistence criterion was a modest but positive correlation. The results sup-

TABLE 2
Meta-Analysis of GMAT and UGPA Validities: Total Sample

	<i>N</i>	<i>k</i>	r_{obs}	SD_{obs}	SD_{res}	ρ	SD_{ρ}	90% cred.
First-year Graduate Grade Point Average								
Verbal	48,915	315	.23	.10	.06	.34	.09	.19–.49
Quantitative	48,758	314	.29	.11	.08	.38	.11	.20–.56
Total	28,624	202	.32	.11	.07	.47	.11	.29–.65
UGPA	50,138	324	.25	.10	.05	.35	.07	.23–.47
Graduate Grade Point Average								
Verbal	5,466	28	.22	.09	.06	.32	.08	.19–.45
Quantitative	5,609	27	.23	.10	.07	.30	.10	.14–.46
Total	5,201	29	.31	.12	.08	.47	.12	.27–.67
UGPA	5,609	28	.25	.12	.09	.35	.12	.15–.55
J-S UGPA	1,292	7	.23	.09	.05	.31	.07	.19–.43
Persistence								
Verbal	680	4	.07	.06	.00	.10	.00	0.1–0.1
Quantitative	680	4	.11	.10	.06	.13	.07	.01–.25
Total	680	4	.12	.09	.04	.17	.05	.09–.25
UGPA	637	3	.09	.14	.12	.11	.15	–.14–.36

Note. *N* = number of subjects; *k* = number of studies; r_{obs} = sample size weighted mean observed correlation; SD_{obs} = observed standard deviation; SD_{res} = residual standard deviation; ρ = operational validity; SD_{ρ} = standard deviation of true validity; 90% cred. = 90% credibility interval.

port this hypothesis. Persistence was predicted by the GMAT-Verbal ($N = 680$, $k = 4$), GMAT-Quantitative ($N = 680$, $k = 4$), GMAT-Total ($N = 680$, $k = 4$), and UGPA ($N = 637$, $k = 3$) with operational validities of .10, .13, .17, and .11, respectively.

Meta-analyses of GMAT and UGPA validities for First-Year GPA were conducted to examine the potential moderating effect of (a) gender, (b) native language of students, (c) full-time versus part-time student status, and (d) undergraduate major (i.e., business major vs. nonbusiness major, and quantitative major vs. nonquantitative major). Results are presented in Table 3. We hypothesized no differences across moderator groups. Overall confidence intervals across moderator groups overlapped except for full- versus part-time students. Validities for the GMAT were somewhat stronger for part-time students while UGPA was a somewhat weaker predictor for part-time students.

The operational validities for females ($N = 576$, $k = 10$) of .28, .39, and .40 for GMAT-Verbal, GMAT-Quantitative, and UGPA, respectively, were comparable to those reported for samples comprised only of males ($N = 1,315$, $k = 10$) of .26, .32, and .32, respectively.

The studies examining the validity of GMAT scores for nonnative English-speaking individuals were not included in the overall analysis and were considered separately. The validity estimates for this group ($N = 1,815$, $k = 20$) were .21 for GMAT-Verbal and .35 for GMAT-Quantitative. The GMAT-Verbal validity is lower than the estimate for native English-speaking students.

Validity estimates for full-time students for GMAT-Verbal ($N = 705$, $k = 7$), GMAT-Quantitative ($N = 541$, $k = 6$), and UGPA ($N = 705$, $k = 7$) were .32, .32, and .37, respectively. For part-time students, the validity estimates for GMAT-Verbal ($N = 735$, $k = 10$), and GMAT-Quantitative ($N = 799$, $k = 10$) were .41 and .39, respectively, while the validity estimate for UGPA ($N = 1,102$, $k = 14$) was .27. Confidence intervals did not overlap for these groups.

The validity estimates for students with business undergraduate degrees and those with nonbusiness undergraduate degrees were almost identical. For students with business undergraduate degrees, the validity estimates for GMAT-Verbal ($N = 1,499$, $k = 15$), GMAT-Quantitative ($N = 1,499$, $k = 15$), and UGPA ($N = 1,163$, $k = 14$) were .33, .32, and .35, respectively. For students with nonbusiness undergraduate degrees, the respective validity estimates were .35 for GMAT-Verbal ($N = 1,298$, $k = 14$), .33 for GMAT-Quantitative ($N = 1,298$, $k = 14$), and .35 for UGPA ($N = 1,298$, $k = 14$).

Largely similar validity estimates were also observed for students with quantitative undergraduate degrees and students with nonquantitative undergraduate degrees. The validity estimates for students with quantitative undergraduate degrees ($N = 483$, $k = 7$) for GMAT-Verbal, GMAT-Quantitative, and UGPA were .37, .40, and .41, respectively. The validity estimates for students with nonquantitative undergraduate degrees ($N = 764$, $k = 7$), were .35, .40, and .39, for GMAT-Verbal, GMAT-Quantitative, and UGPA, respectively.

TABLE 3
Meta Analyses of GMAT Validity for First-year GPA Across Different Groups

	<i>N</i>	<i>k</i>	<i>r</i> _{obs}	<i>SD</i> _{obs}	<i>SD</i> _{res}	ρ	<i>SD</i> _{ρ}	90% cred.
Gender								
Female - GMAT-V	576	10	.19	.19	.13	.28	.20	-.05-.61
Female - GMAT-Q	576	10	.30	.20	.15	.39	.20	.06-.72
Female - UGPA	576	10	.29	.17	.11	.40	.14	.17-.63
Male - GMAT-V	1,315	10	.18	.14	.10	.26	.15	.01-.51
Male - GMAT-Q	1,315	10	.24	.11	.07	.32	.09	.17-.47
Male - UGPA	1,315	10	.23	.08	.00	.32	.00	.32-.32
Nonnative English-speaking Students (NNES)								
NNES - GMAT-V	1,815	20	.14	.10	.00	.21	.00	
NNES - GMAT-Q	1,815	20	.26	.12	.07	.35	.08	
Full-Time vs. Part-Time Students								
Full-Time - GMAT-V	705	7	.22	.07	.00	.32	.00	.32-.32
Full-Time - GMAT-Q	541	6	.24	.06	.00	.32	.00	.32-.32
Full-Time - UGPA	705	7	.27	.06	.00	.37	.00	.37-.37
Part-Time - GMAT-V	735	10	.28	.09	.00	.41	.00	.41-.41
Part-Time - GMAT-Q	799	10	.30	.18	.14	.39	.18	.09-.69
Part-Time - UGPA	1,102	14	.20	.12	.03	.27	.05	.19-.35
Business Undergraduate Degree vs. Nonbusiness Undergraduate Degree								
Business - GMAT-V	1,499	15	.23	.15	.11	.33	.16	.07-.59
Business - GMAT-Q	1,499	15	.25	.16	.12	.32	.16	.06-.58
Business - UGPA	1,163	14	.25	.07	.00	.35	.00	.35-.35
Non-Business - GMAT-V	1,298	14	.24	.11	.05	.35	.07	.23-.47
Non-Business - GMAT-Q	1,298	14	.25	.13	.08	.33	.11	.15-.51
Non-Business - UGPA	1,298	14	.25	.06	.00	.35	.00	.35-.35
Quantitative Undergraduate vs. Nonquantitative Undergraduate								
Quantitative - GMAT-V	483	7	.25	.12	.01	.37	.02	.34-.40
Quantitative - GMAT-Q	483	7	.31	.14	.09	.40	.12	.20-.60
Quantitative - UGPA	483	7	.30	.11	.00	.41	.00	.41-.41
Nonquant. - GMAT-V	764	7	.24	.10	.03	.35	.05	.27-.43
Nonquant. - GMAT-Q	764	7	.31	.11	.06	.40	.08	.27-.53
Nonquant. - UGPA	764	7	.28	.06	.00	.39	.00	.39-.39

Note. *N* = number of subjects; *k* = number of studies; *r*_{obs} = sample size weighted mean observed correlation; *SD*_{obs} = observed standard deviation; *SD*_{res} = residual standard deviation; ρ = operational validity; *SD* _{ρ} = standard deviation of true validity; 90% cred. = 90% credibility interval.

DISCUSSION

Our study found considerable support for the validity of the GMAT. Across all criteria and moderator groups examined, the results indicate that the GMAT is predictive of business school performance. Not only do the GMAT and UGPA predict first-year grades, but they also correlate with overall graduate record. Results also indicate that the GMAT does predict, although modestly, persistence in business school. Results for nonnative English-speaking students are consistent with previous research, with lower positive validities for the Verbal measures and moderately high validities for the Quantitative measures. Finally, within the limits of the available data, the evidence we obtained suggests the GMAT is not strongly moderated by gender or academic background variables with the exception of small moderator effect for full- versus part-time status, where the GMAT was more valid for part-time students, and UGPA was

more valid for full-time students. Part of this difference may be due to predictor variability differences across groups (which could not be directly addressed due to limited information in the literature) or differences in grading and evaluation of students in different programs. For the other four moderator groups, confidence intervals calculated using conservative formulas in Hunter and Schmidt (2004) yielded substantially overlapping intervals across all moderator pairs (e.g., Male GMAT Verbal versus Female GMAT Verbal), suggesting no reliable validity differences.

These findings are important, because they indicate that using the GMAT does indeed have utility for selecting students into graduate schools of business. It is particularly interesting to note that the combined use of the Verbal and Quantitative scales of the GMAT appear to be superior to undergraduate grade point average in predicting graduate student performance. Although consis-

tent with previous research, this observation still requires an explanation. One is that the nature of the work to be done in graduate school is sufficiently different from that in college that more general assessments of ability become more important. An alternative explanation is that the specific content of the GMAT is a better match with the knowledge and skills needed in business school than are the individual difference characteristics that are assessed by undergraduate record. Finally, the validity of UGPA may be diluted by content from courses that are largely unrelated to graduate studies. This explanation is not well supported by the data, as the validity of UGPA for business undergraduate majors is no better than is UGPA for nonbusiness majors. Nonetheless, the results suggest that the best approach for admitting students is the combination of GMAT and UGPA data.

Despite the validity information, a series of other issues need to be addressed surrounding the use of the GMAT in admissions. The first is the extent to which the effects presented here are of meaningful magnitude. The second is the evidence for or against predictor bias. Finally, it is important to consider other predictors that may improve the quality of admissions.

Predictive Power

A common criticism directed toward predictors of performance is that the observed correlations are too low. For example, Melnick (1971) described a validity estimate of .46 as a "weak" predictor that "approaches respectability," when such a correlation is actually far from a weak relationship and has a great deal of applied value. Such reactions probably stem from the confusions about the connection between correlations and their impact on admission decisions. Reframing these effects in terms of accurate versus inaccurate decisions can be quite instructive (Taylor & Russell, 1939). The extent to which a set of admissions tools is sufficiently useful is ultimately a local decision based on their analysis of gains; however, the gains from the GMAT strike us as important in most scenarios.

For example, if we assume that across all applicants to a business school, 70% would be successful by some criterion (say, a "B" average or better), then using no admission system whatsoever will, of course, result in a 30% failure rate and a 70% success rate (30% below B average, 70% above a B average). In the case of a highly selective business school (one that admits 20% of applicants), using the GMAT-Total with a validity of .47 for predicting

grades would drop the failure rate from 30% to 10%. We feel that obtaining a 20% improvement from a few hours of testing can hardly be called small, weak, or trivial. Even the modest correlations between GMAT and persistence can be of use. Assuming that 80% of students will persist through a selective business program, we can reduce the number of failures by 9% by using GMAT-Total. Even with effects only half as large as the estimates described here, such improvements in decision making will have important implications for the success of the student body and the success of graduate business programs.

Bias and Fairness in Admissions Predictors

Of great concern in admissions is that all predictors be unbiased. Considerable research has examined the issue of predictor bias in general and test bias in particular. The most common approach to examining test bias is to use multiple regression to determine if the relationships between scores and performance in school is the same across different groups. Any given score on the predictor should be associated with the same level of performance in school regardless of group membership. Put simply, any score on the GMAT should be associated with the same level of success in school regardless of student gender, race, or ethnicity. This is evaluated by testing to see if the slopes and intercepts for different groups are the same. Recent research by Sireci and Talento-Miller (in press) has found this to be true for the GMAT. These findings are consistent with research on other tests (e.g., Breland, 1979; Pennock-Roman, 1990; Young, 1991). Campbell (1996: 134) summarized the literature, stating, "The preponderance of evidence supports the conclusion that a particular score means the same thing, in terms of what it predicts, for both groups."

Although this research suggests that the GMAT is not unfairly treating test-takers, it does not alter concerns about diversity in admissions. Diversity is a goal for many institutions and use of admissions measures such as the GMAT or UGPA do not prevent programs from pursuing diversity goals. Universities need to specify what cultural, racial, gender, or other background characteristics they feel will positively contribute to the education of their students while meeting legal requirements, and then incorporate them into the admissions decision in the manner and to the extent permissible by law.

Value Added and Gate Keeping

One ongoing concern with the MBA is that it adds value rather than functions as a gatekeeper. As outsiders to this debate, we are, frankly, puzzled by the angst, hair tearing, and chest beating over this issue. Perhaps some of the difference is that most fields content themselves with conveying useful knowledge to students without any expectation that they will produce socially responsible, ethically centered, globally aware, charismatic leaders with deep characters and profound reasoning skills. All, we might add, in an 18–24-month mostly classroom-based intervention. We feel that a simple evaluation of the core content of MBA programs quickly answers the question of value added. If the content, exercises, and evaluations target information that is worth learning, then there is value added. For professional work in business, topics such as economics, finance, accounting, organizational behavior, human resources, international business, quantitative analysis, strategy, operations, and, of course, ethics are all worth learning.

The notion of gate keeping seems to further break down from another perspective. Nearly any college graduate can obtain admission to an MBA program, as the majority of programs are not highly selective. Based on our analysis of selectivity data, it is the rare program that admits less than half of all applicants (GMAC, 2002). When we avert our gaze from the brand-name high-prestige MBA programs, we see that the majority of MBA matriculants did not have to leap over a tall building to gain admission to a program and were at little risk of failing out while they were there. It is certainly true that the effect of MBA program prestige on success is an interesting topic for study. However, the MBA itself is actually far less of a gatekeeper than is often portrayed.

The GMAT as a Training Tool

One possible use for an assessment is to identify the specific strengths and weaknesses of an applicant. This information about training needs can then be used to develop targeted training or educational interventions (Campbell & Kuncel, 2001). The GMAT does provide global information about verbal, writing, and quantitative skills. As such, this information could be used to identify general training needs for the student. Poor analytic writing could be addressed with extra writing courses, and weak quantitative skills can be shorn up with math and statistics education.

However, the focus of the GMAT is intentionally very broad in content and to assess more specific

training needs would require a battery of very specific knowledge and skill measures or rating data ideally from multiple sources. This approach to training and development is somewhat common in organizations but is all but unheard of in education outside of basic course requirements. It need not be. A behavior-based model of student performance would need to be carefully developed. This model would inform assessments and feedback, which could be accomplished by an evaluation of specific course requirements combined with peer and faculty evaluations of students (i.e., multirater feedback). Ideally, the model would also be linked to those skills that organizations value from new MBAs. This approach could be very expensive. An easier and less expensive method would be to use prior academic record and a structured recommendation form for identifying student-training needs. Both could be done to provide ongoing feedback and customization of a student's training. However, this really goes beyond the purpose of the GMAT, which focuses on readiness to learn.

Although a more elaborate assessment system is appealing, it should be used for postadmission feedback rather than as a preadmission selection tool. At its heart, the MBA is typically a generalist degree. In contrast, many domains assume that applicants already have considerable discipline-specific knowledge before entering. Discipline-specific admission measures (e.g., the GRE Subject Tests) make sense for these degree programs. For many MBA programs, however, the goal is to provide broad business training that is applicable across many jobs in many organizations. A broader evaluation of verbal and quantitative skills is appropriate in this context. The extent to which evaluations could be more tightly focused on the tasks of business without including considerable business-specific knowledge would need to be assessed. Only highly specialized programs where the focus is on exhaustive knowledge of narrow domains would likely benefit from more specific assessments of specific business knowledge. In addition, specific knowledge assessments would tend to filter out many students who do not have prior training in related fields. The desirability of these effects would need to be evaluated by specific programs.

Alternate Predictors of Student Performance

Our findings here do not mean that other characteristics are unimportant for success in graduate school or business. The goals of diversity and improved prediction have driven researchers to examine other possible predictors. The features im-

portant for the GMAT and UGPA also apply to these predictors. In particular the measures need to be highly valid, low in cost, contribute to existing measures, and not yield adverse impact.

Research on predictors of job performance has experienced something of a research renaissance in the area of personality predictors of work performance (e.g., Hough & Ones, 2001; Roberts & Hogan, 2001). We have every reason to expect that these traits will be related to academic performance as well. In fact, research on business school programs suggests that noncognitive variables have important relationships with performance (Rothstein, Paunonen, Rush, & King, 1994).

Situational judgment and case study measures have recently been developed with the goal of improving admissions decisions and lowering adverse impact in MBA admissions (Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2005). So far the results are quite mixed. The measures provide nominal incremental validity over the GMAT and UGPA with individual scales increasing Multiple-*R* in a range from .01 to .04. Unfortunately, the measures demonstrate smaller predictive validities than the GMAT or UGPA. In addition, these measures also yield racial score differences. Fortunately, some of the lower validity is likely to be due to poor reliability, which could be addressed through more measurement development. However, increased reliability is a double-edged sword in this case. It should yield increased predictive validity but will also yield increased group differences and adverse impact. What is most alarming is that the Case Study measure, unlike standardized tests, yields group differences favoring Whites for *all* examined racial groups, including Asians. Unfortunately, it is also the Case Study measure that has the strongest predictive and incremental validities. Finally, it is unknown at this time if the new measures yield test bias, as the current samples reported in the literature are of insufficient size to conduct such analyses.

A final big-picture issue with most alternative assessments is that they require self-reports or judgments on relatively transparent problems. Situational judgment tests, biodata measures, and personality measures are all subject to faking and response distortion. Faking on a situational judgment measure appears to yield a large increase in scores (Peeters & Lievens, 2005). It is important to note that this increase, without any extensive coaching, is far *greater* than the effect of extensive formal test preparation on standardized tests (e.g., Powers & Rock, 1998). Future researchers need to examine the predictive power of these measures under more realistic operational situations. An al-

ternative would be to assess these characteristics through ratings provided by people other than the student (e.g., faculty, academic advisors). In summary, a great deal of additional research and validation work is needed to see if these measures can be useful, and even if found to be useful their current role appears to be limited to being a supplement to the GMAT rather than a replacement.

Limitations and Directions for Future Research

This study represents a statistical synthesis of the available literature and the results suggest that differences across schools and cohorts have a limited effect on the validity of the GMAT. Like all studies the generalizability of the results presented here are constrained by the available data. The data obtained indicate that the validity of the GMAT is quite likely to generalize across situations. The data supporting this conclusion comes from many different institutions across a wide timespan as reflected in the literature. However the data are not a random sample of all situations. As such the validity of the GMAT may be moderated by situations unexamined by this study. Similarly, the moderator analyses are also constrained by the available data. Although analyses are still based on hundreds of individuals, additional research is needed to more fully examine moderator effects. This possibility may motivate universities to conduct local validation studies. Such studies are clearly important but should be regarded with caution given the relatively small amount of information contained in an individual study relative to a large meta-analysis. In a local validation study it is critical to consider range restriction on the GMAT and UGPA, quality of grading standards, and the variance of grades in business school. If a program consistently does a poor job awarding grades or awards high grades to nearly all students, then there is little to predict. Finally, this study examined an important, but narrow set of measures of student performance. Expanding the criterion space will be important for future work on MBA student performance.

Developing multidimensional models of performance in business school will be important for enhancing our understanding of all of the effective and ineffective behaviors exhibited in business school and will aid the development of additional predictors. Work by Hilton, Kendall, and Sprecher (1970) attempted to specify important areas of performance for business students. Much more of this type of work is needed, as there are likely to be important aspects of performance that are not currently considered. Identifying these behaviors can

only improve our ability to select and train the best students. It can also improve our ability to understand how behaviors exhibited in business school are related to professional job performance.

This meta-analysis supports the validity of the GMAT for predicting grades at earlier and later stages in business school and persisting through the program across the general population of students. In addition the combination of GMAT scales is a better predictor than prior grades with the combination of GMAT and UGPA likely yielding the best predictive validity. Therefore, we should not expect that the use of the GMAT for selecting business school students would result in a limited management workforce. We should expect that it instead would enhance the quality of business school admissions and the workforce. Arguments that the GMAT is not a valid predictor of student performance in graduate business school are in error.

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.
- Aitken, A. C. 1934. Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society*, 4: 106-110.
- Barritt, L. S. 1966. The consistency of first-semester college grade point average. *Journal of Educational Measurement*, 3: 261-262.
- Bendig, A. W. 1953. The reliability of letter grades. *Educational and Psychological Measurement*, 13: 311-321.
- Bieker, R. F. 1996. Factors affecting academic achievement in graduate management education. *Journal of Education for Business*, 72: 42-46.
- Bottger, P., & Yetton, P. 1982. Student assessment and GMAT: Quantitative versus verbal components in performance on examination and assignments. *Australian Journal of Management*, 7: 9-18.
- Breland, H. 1979. *Population validity and college entrance measures. (Research Monograph No. 8)*. New York: The College Entrance Examination Board.
- Campbell, J. P. 1996. Group differences and personnel decisions: Validity, fairness, and affirmative action. *Journal of Vocational Behavior*, 49, 122-158.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E., Jr. 1970. *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.
- 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*. San Francisco: Jossey-Bass.
- Campbell, J. P., & Kuncel, N. R. 2001. Individual and team training. In N. Anderson, D. S. Ones, H. Sinangil, & C. Viswesvaran, (Eds.), *Handbook of industrial, work, and organizational psychology*: 287. London, UK: Sage.
- Cascio, W. F. 1998. *Applied psychology in personnel management*, 5th Edition. Upper Saddle River, NJ: Prentice Hall.
- Dobson, P., Krapljan-Barr, P., & Vielba, C. 1999. An evaluation of the validity and fairness of the Graduate Management Admissions Test (GMAT) used for MBA selection in a UK business school. *International Journal of Selection & Assessment*, 7: 196-202.
- Dreher, G. F., & Ryan, K. C. 2000. Prior work experience and academic achievement among first-year MBA students. *Research in Higher Education*, 41: 504-525.
- Dreher, G. F., & Ryan, K. C. 2004. A suspect MBA selection model: The case against the standard work experience requirement. *Academy of Management Learning and Education*, 3: 87-91.
- Educational Testing Service. 2003. *College-bound seniors: A profile of SAT program test takers 2003*. Educational Testing Service. Princeton, NJ: Author.
- Enright, M. K., & Gitomer, D. 1989. *Toward a description of successful graduate students*. GRE Board Report No. 85-17R. Princeton, NJ: Educational Testing Service.
- Fairtest. 2003. *GMAT- Padlock on MBA Admissions Gates*. <http://www.fairtest.org/facts/GMAT%20Fact.html> Accessed September 16, 2003.
- Goodrich, J. N. 1975. American standardized tests: Pseudo-indicators of ability? *Educational Technology*, 15: 23-25.
- Graduate Management Admission Council. 2002. Unpublished raw data.
- Graduate Management Admission Council. 2003. *Profile of Graduate Management Admission Test Candidates, 1997-1998 to 2001-2002*. Graduate Management Admission Council. McLean, VA: Author.
- Grambsch, P. V. 1981. Business Administration. In A. W. Chickering (Ed.), *The modern American college*: 472-486. New York: Jossey-Bass.
- Hancock, T. 1999. The gender difference: Validity of standardized admission tests in predicting MBA performance. *Journal of Education for Business*, 75: 91-93.
- Hedlund, J., Wilt, J. M., Nebel, K. R., Ashford, S. J., & Sternberg, R. J. 2005. *Assessing practical intelligence in business school admissions: A supplement to the Graduate Management Admissions Test*. Unpublished manuscript.
- Hilton, T. L., Kendall, L. M., & Sprecher, T. B. 1970. *Performance criteria in graduate business study. Parts I and II - Development of rating scales, background data form and the pilot study*. Research Bulletin 70-3. Educational Testing Service. Princeton, NJ.
- 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*, 1: 233-277. London, UK: Sage.
- Hunter, J. E., & Schmidt F. L. 2004. *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Humphreys, L. G. 1989. Intelligence: Three kinds of instability and their consequences for policy. In R. L. Linn (Ed.), *Intel-*

- ligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys:**193–216. Champaign, IL: University of Illinois Press.
- Humphreys, L. G., Taber, T. 1973. Postdiction study of the Graduate Record Examination and eight semesters of college grades. *Journal of Educational Measurement*, 10, 179–184.
- Humphreys, L. G. 1968. The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology*, 59: 375–380.
- Kuncel, N. R. 2003. *The prediction and structure of academic performance*. Unpublished doctoral dissertation. Minneapolis, MN: University of Minnesota.
- Kuncel, N. R., Crede, M., Thomas, L. L., Klieger, D. M., Seiler, S. N., & Woo, S. E. 2005. A meta-analysis of the validity of the pharmacy college admission test (PCAT) and grade predictors of pharmacy student performance. *American Journal of Pharmaceutical Education*, 69: 339–347
- 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.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. 2004. Predicting academic performance, job performance, career potential, and creativity: Can one construct do it all? *Journal of Personality and Social Psychology: Special Issue, Cognitive Abilities: 100 Years after Spearman (1904)*, 68: 148–161.
- Lawley, D. N. 1943. A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, LXII-Part I(I)*: 28–30.
- 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, (Eds.), *Learning and individual differences: Process, trait, and content determinants*: 55–76. Iowa City, IA: University of Iowa, College of Education, Division of Psychological & Quantitative Foundation.
- 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.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. 1988. Job experience correlates of job performance. *Journal of Applied Psychology*, 73: 327–330.
- McKelpin, J. P. 1965. Some implications of the intellectual characteristics of freshmen entering a liberal arts college. *Journal of Educational Measurement*, 2: 161–166.
- Melnick, M. 1971. *Predicting graduate business grades with the admission test for graduate study in business*. Brief Report No. 25, Hofstra University, Hempstead, NY.
- Melnick, M. 1972. *A further investigation of predictors of success in the MBA program at Hofstra University*. Report No. R-98, Hofstra University, Hempstead, NY.
- Michell, K. J. 1990. Traditional predictors of performance in medical school. *Academic Medicine*, 65: 149–158.
- Morris, S. S. 1995. *The relationship between self-directed learning readiness and academic performance in a nontraditional higher education program*. Unpublished doctoral dissertation, The University of Oklahoma, Norman.
- Peeters, H., & Lievens, F. 2005. Situation judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65: 70–89.
- Pennock-Roman, M. 1990. *Test validity and language background: A study of Hispanic American students at six universities*. New York: The College Board.
- Powers, D. E., & Rock, D. A. 1998. Effects of coaching on SAT I: Reasoning Test scores. *Journal of Educational Measurement*, 36: 93–118.
- Ree, M. J., Carretta, T. R., & Earles, J. A. 1998. In top-down decisions, weighting variables does not matter: A consequence of Wilks' theorem. *Organizational Research Methods*, 1: 407–420.
- 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. Clifford (Eds.), *Policy issues in employment testing*: 131–224. Boston: Kluwer.
- Roberts, B. W., & Hogan, R. (Eds.). 2001. *Personality psychology in the workplace*. Washington, DC: American Psychological Association.
- Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. 1994. Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, 86: 516–530.
- Sackett, P. R., & Yang, H. 2000. Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, 85: 112–118.
- 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., & Hunter, J. E. 1980. Validity generalization results for computer programmers. *Journal of Applied Psychology*, 65: 643–661.
- Sireci, S. G., & Talento-Miller, E. In press. Evaluating the predictive validity of Graduate Management Admission Test Scores. *Educational and Psychological Measurement*.
- Shuey, A. M. 1950. Choice of major subject as related to American Council Examination score and college grades. *Journal of Educational Psychology*, 41: 292–300.
- Stanely, J. D., & Porter, A. C. 1967. Correlation of Scholastic Aptitude Test Scores with college grades for Negroes versus Whites. *Journal of Educational Measurement*, 4: 199–218.
- Sternberg, R. J., & Wagner, R. K. 1993. The g-centric 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 effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23: 565–578.
- Thorndike, R. L. 1949. *Personnel selection: Test and measurement techniques*. New York: Wiley.
- Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. 2001. *Selecting successful graduate students: In-depth interviews with GRE users*. GRE Board Report. Princeton, NJ: Educational Testing Service.

- Wernimont, P. F., & Campbell, J. P. 1968. Signs, samples, and criteria. *Journal of Applied Psychology*, 52: 372-376.
- Young, J. E. 1991. Gender bias in predicting college academic performance: A new approach using Item Response Theory. *Journal of Educational Measurement*, 28: 37-47.
- Zwick, R. 2002. *Fair game? The use of standardized admissions tests in higher education*. New York: Routledge-Falmer.
- ### Studies Included in the Meta-Analysis
- Adams, A. J., & Hancock, T. 2000. Work experience as a predictor of MBA performance. *College Student Journal*, 34: 211-217.
- Ahmadi, M., Raiszadeh, F., & Helms, M. 1997. An examination of the admission criteria for the MBA program: A case study. *Education*, 117: 540-546.
- Arnold, L. R., & Balakrishnan, C. 1996. Applicant evaluation in an executive MBA program. *Journal of Education for Business*, 71: 277-283.
- Baird, L. L. 1975. Comparative prediction of first year graduate and professional school grades in six fields. *Educational & Psychological Measurement*, 35: 941-946.
- Benson, G. 1983. *GMAT—fact or fiction: A look at the validity of the exam*. Paper presented at the Annual Meeting of the Rocky Mountain Educational Research Association, Tucson, AZ.
- Bottger, P., & Yetton, P. 1982. Student assessment and GMAT: Quantitative versus verbal components in performance on examination and assignments. *Australian Journal of Management*, 7: 9-18.
- Braunstein, A. W. 2002. Factors determining success in a graduate business program. *College Student Journal*, 36: 471-477.
- Carver, M. R., Jr., & King, T. E. 1994. An empirical investigation of the MBA admission criteria for nontraditional programs. *Journal of Education for Business*, 70: 94-98.
- Daft, R. L. 1979. *MBA admission criteria, communication skill, and academic success: An unexpected finding*. Paper presented at the thirty-ninth annual meeting of the Academy of Management Proceedings, Atlanta, GA.
- Davis, L. F. 1987. *Moral judgment development of graduate management students in two cultures: Minnesota and Singapore*. Unpublished doctoral dissertation, University of Minnesota, St. Paul.
- Dobson, P., Krapljan-Barr, P., & Vielba, C. 1999. An evaluation of the validity and fairness of the Graduate Management Admissions Test (GMAT) used for MBA selection in a UK business school. *International Journal of Selection and Assessment*, 7: 196-202.
- Eber, J. E. 1983. *Pre-admission factors influencing academic success for part-time students in the Master of Business Administration programs of three selected private schools*. Unpublished doctoral dissertation, Northern Illinois University, DeKalb.
- Elkin, G. 1991. Decision-making processes for admitting students to executive MBA programmes. *Management Education and Development*, 22: 134-142.
- Fisher, J. B., & Resnick, D. A. 1990. Standardized testing and graduate business school admission: A review of issues and an analysis of a Baruch College MBA cohort. *College and University*, 65: 137-148.
- Geisler, M. P. 1983. *The older graduate student: A descriptive study*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- Graham, L. D. 1991. Predicting academic success of students in a Master of Business Administration program. *Educational and Psychological Measurement*, 51: 721-727.
- Hancock, T. 1999. The gender difference: Validity of standardized admission tests in predicting MBA performance. *Journal of Education for Business*, 75: 91-93.
- Harrell, T. W., & Alpert, B. 1989. Attributes of successful MBAs: A 20-year longitudinal study. *Human Performance*, 2: 301-322.
- Hecht, L. W., & Schrader, W. B. 1986. *Graduate management admission test*. Graduate Management Admission Council, McLean, VA.
- Hecht, L. W., & Powers, D. E. 1982. *The predictive ability of preadmission measures in graduate management education: Three years of the GMAC validity study service*. Report No. GMAC-RR-82-1. Graduate Management Admission Council. Princeton, NJ.
- Hoefler, P., & Gould, J. 2000. Assessment of admission criteria for predicting students' academic performance in graduate business programs. *Journal of Education for Business*, 75: 225-229.
- Jackson, P. H., & Novick, M. R. 1969. *Maximizing the validity of a unit weight composite as a function of relative component lengths with a fixed total testing time*. Research Bulletin 69-1-16. Educational Testing Service Research Bulletin, Princeton, NJ.
- Kaiser, P. R. 1978. Predictors of academic success for MBA students at Indiana University. Unpublished doctoral dissertation, Indiana University, Bloomington.
- Kurtzberg, T. R. 1998. Creative thinking, cognitive aptitude, and integrative joint gain: A study of negotiator creativity. *Creativity Research Journal*, 11: 283-293.
- Lee, E. B. 1977. *Communication and cognition: Differences among students in journalism, law, and business*. Unpublished doctoral dissertation, University of Texas, Austin.
- Lewis, J. W. 1964. Caution about using test composites as predictors. *Journal of College Student Personnel*, 6: 95-97.
- Lewis, J. W. 1964. The relationship of selected variables to achievement and persistence in a masters program in business administration. *Educational and Psychological Measurement*, 24: 951-954.
- Malliaris, M. E. 1981. *The use of multiple regression residuals analysis in restructuring prediction equations*. Unpublished doctoral dissertation, Loyola University, Chicago.
- Marks, R., Watt, P., & Yetton, P. 1981. GMAT scores and performance: Selecting students into a graduate management school. *Australian Journal of Management*, 6: 81-102.
- Melnick, M. 1972. *A further investigation of predictors of success in the MBA program at Hofstra University*. Report No. R-98, Center for the Study of Higher Education. Hempstead, NY.
- Morris, S. S. 1995. *The relationship between self-directed learning readiness and academic performance in a nontraditional higher education program*. Unpublished doctoral dissertation, University of Oklahoma, Norman.
- Nilsson, J. E. 1995. The GRE and the GMAT: A comparison of their correlations to GGPA. *Educational and Psychological Measurement*, 55: 637-640.
- Novick, M. R., & Thayer, D. T. 1969. *Some applications of the*

- procedures for allocating testing time.* Research Bulletin RB-69-1, Educational Testing Service, Princeton, NJ.
- Olsen, M. 1957. *The admission test for graduate study in business as a predictor of first year grades in business schools, 1954-1955.* Statistical Report SR-57.3, Educational Testing Service, Princeton, NJ.
- Paolillo, J. G. P. 1982. The predictive validity of selected admissions variables relative to grade point average earned in a master of business administration program. *Educational and Psychological Measurement*, 42: 1163-1167.
- Peiperl, M. A., & Trevelyan, R. 1997. Predictors of performance at business school and beyond. Demographic factors and the contrast between individual and group outcomes. *Journal of Management Development*, 16: 354-367.
- Pfeffer, J. 1977. Effects of an MBA and socioeconomic origins on business school graduates' salaries. *Journal of Applied Psychology*, 62: 698-705.
- Pitcher, B. 1972. *Report of validity studies carried out by ETS for graduate schools of business 1954-1970.* Educational Testing Service, Princeton, NJ.
- Powers, D. E. 1980. *The relationship between scores on the Graduate Management Admission Test and the Test of English as a Foreign Language.* Research Report ETS-RR-80-31, TOEFL-RR-5. Educational Testing Service, Princeton, NJ.
- Powers, D. E., & Evans, F. R. 1978. *Relationships of preadmission measures to academic success in graduate management education.* Educational Testing Service, Princeton, NJ.
- Powers, D. E., & Moss, P. A. 1980. *A summary of the results of the Graduate Management Admissions Council validity study service for 1979-1980.* Report No. RR-80-29, Educational Testing Service, Princeton, NJ.
- Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. 1994. Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, 86: 516-530.
- Schwan, E. S. 1988. MBA admission criteria: An empirical investigation and validation study. *Journal of Education for Business*, 63: 158-162.
- Tanchareonrat, O. 1988. *Selected characteristics and academic achievement of international graduate students at Illinois State University.* Unpublished doctoral dissertation, Illinois State University, Normal.
- Waters, E., Jr. 1968. *The Admission Test for Graduate Study in Business and the adult part-time student.* George Washington University, Washington, DC.
- Wilson, K. M. 1985. *Factors affecting GMAT predictive validity for foreign MBA Students: An exploratory study.* Report No. ETS-RR-85-17, Educational Testing Service, Princeton, NJ.
- Wright, R. E., & Palmer, J. C. 1997. Examining performance predictors for differentially successful MBA students. *College Student Journal*, 31: 276-281.
- Wright, R. E., & Palmer, J. C. 1998. Predicting performance of above and below average performers in graduate business schools: A split sample regression analysis. *Educational Research Quarterly*, 22: 72-79.
- Youngblood, S. A., & Martin, B. J. 1982. Ability testing and graduate admissions: Decision process modeling and validation. *Educational and Psychological Measurement*, 42: 1153-1162.

This study was funded by the Graduate Management Admissions Council by a grant to Nathan R. Kuncel. Nathan Kuncel received no personal compensation for any work on the study. The opinions are those of the authors and not necessarily those of GMAC.

Nathan R. Kuncel received his doctorate in industrial and organizational psychology from the University of Minnesota. He was delighted to return to his alma mater as an assistant professor of psychology in 2005. Dr. Kuncel's research focuses on the prediction and structure of performance in academic and work settings.

Marcus Credé received his doctorate in industrial and organizational psychology from the University of Illinois-Urbana Champaign and is currently faculty at Fairleigh Dickenson University. Dr. Crede's research focuses on job attitudes and personnel selection.

Lisa L. Thomas, MA, is currently pursuing a doctoral degree at the University of Illinois-Urbana Champaign. Thomas' research interests include (but are not limited to) selection and assessment, psychological testing, psychological measurement, dynamic criteria, and the prediction of performance.

Copyright of *Academy of Management Learning & Education* is the property of *Academy of Management* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.