

Validity of the Medical College Admission Test for Predicting Medical School Performance

Ellen R. Julian, PhD

Abstract

Purpose

Since the introduction of the revised Medical College Admission Test (MCAT®) in 1991, the Association of American Medical Colleges has been investigating the extent to which MCAT scores supplement the power of undergraduate grade point averages (uGPAs) to predict success in medical school. This report is a comprehensive summary of the relationships between MCAT scores and (1) medical school grades, (2) United States Medical Licensing Examination (USMLE) Step scores, and (3) academic distinction or difficulty.

Method

This study followed two cohorts from entrance to medical school through

residency. Students from 14 medical schools' 1992 and 1993 entering classes provided data for predicting medical school grades and academic difficulty/distinction, while their peers from all of the U.S. medical schools were used to predict performance on USMLE Steps 1, 2, and 3. Regression analyses assessed the predictive power of combinations of uGPAs, MCAT scores, and undergraduate-institution selectivity.

Results

Grades were best predicted by a combination of MCAT scores and uGPAs, with MCAT scores providing a substantial increment over uGPAs. MCAT scores were better predictors of USMLE Step scores than were uGPAs, and the combination did little better than MCAT

scores alone. The probability of experiencing academic difficulty or distinction tended to vary with MCAT scores. MCAT scores were strong predictors of scores for all three Step examinations, particularly Step 1.

Conclusions

MCAT scores almost double the proportion of variance in medical school grades explained by uGPAs, and essentially replace the need for uGPAs in their impressive prediction of Step scores. The MCAT performs well as an indicator of academic preparation for medical school, independent of the school-specific handicaps of uGPAs.

Acad Med. 2005; 80:910-917.

Since the introduction of the revised Medical College Admission Test (MCAT®) in 1991, the Association of American Medical Colleges (AAMC) has been investigating the extent to which MCAT scores predict success in medical school and how that supplements the predictive power of undergraduate (i.e., bachelor's degree) grade point averages (uGPAs). During the course of the longitudinal study that is the focus of this report, MCAT staff members have published reports on subsets of the data.¹⁻³ This final report summarizes the previous findings and expands on them with an additional cohort and later outcome measures.

Results from analyses predicting both medical school grades in the first three years and also Step1 scores of the United States Medical Licensing Examination

(USMLE) for the 1992 entering class, or cohort, were reported by Mitchell et al.¹ in 1994 and by Wiley and Koenig² in 1996. The authors of the 1994 report concluded that uGPA, MCAT scores, and school selectivity were all useful in the prediction of first-year performance.

The authors of the 1996 report went a step further and examined the extent to which the previously used predictor variables were predictive of first- and second-year medical school performance and of USMLE Step 1 scores. They found that MCAT scores had slightly higher correlations with medical school grades than did uGPAs, and noted that the validity coefficients all improved when MCAT and uGPA data were both considered.

Koenig et al.³ used some of the same data in a study of how well the MCAT predicted performance for different races and sexes. That study was a direct response to claims that the MCAT was biased against women and minorities and sought to determine if race and sex needed to be predictor variables in future prediction equations. The investigators

used the total score on the three multiple-choice sections of the MCAT, uGPAs, school selectivity, sex, and race to predict USMLE Step 1 scores. They found that uGPAs and MCAT scores were useful predictors of Step1 scores for all groups, and that there was no evidence of underprediction (i.e., bias) for either women or the minority groups. As a result, race and sex were not included in subsequent prediction equations.

Hojat et al.⁴ examined the predictive power of the MCAT Writing Sample. Using data for matriculating students at Jefferson Medical College, the researchers divided matriculants into top, middle, and bottom groups based on their Writing Sample scores. They found that, while the top group outperformed the bottom group in first- and second-year basic science courses and on Step 1, the differences were not statistically significant. They found, however, statistically significant differences among Writing Sample groups' scores on a number of performance measures in clinical disciplines.

Dr. Julian is associate vice president and director, Medical College Admission Test, Association of American Medical Colleges, Washington, D.C.

Correspondence should be addressed to Dr. Julian, Association of American Medical Colleges, 2450 N Street, N.W., Washington, D.C. 20037-1127.

Numerous studies have evaluated the predictive validity of other standardized tests. Recently, Bridgeman et al.⁵ used SAT scores, high school curriculum intensity, and high-school GPA to predict academic achievement in college. Previous, regression-based research had typically shown that SAT scores added little to prediction (less than 10% of additional variance explained). Bridgeman et al., however, used a performance-groups comparison method and found that SAT scores did contribute significantly to the prediction of both college freshman year and four-year uGPA (i.e., the average uGPA over the four years of college), especially for students who did very well in high school.

Kuncel et al.⁶ performed a massive meta-analysis of 1,521 studies with 6,589 correlations relating to the use of Graduate Record Examination (GRE) scores, for both general and subject examinations, to predict graduate school outcomes. Of most applicability to the present report is their analysis of the power of GRE quantitative, verbal, and subject examinations to predict graduate GPAs. Unlike the research reported in this article, their emphasis was on the incremental validity of the uGPA added to GRE scores, rather than the other way around. They found that the GRE scores alone generated an expected validity coefficient, corrected for restriction in range, of .50 (25% of the variance explained), and the addition of the uGPA brought the coefficient up to .54 (29% of the variance). Those authors also point out that the expected utility of algorithmic combinations of predictor

data for selection decisions is substantially attenuated by admission committees' common use of subjective and variable weightings of the variables.

The present report is a comprehensive summary of the relationships between uGPAs and MCAT scores and (1) medical school grades, (2) USMLE Step scores, and (3) academic distinction or difficulty.

Method

MCAT scores are not used in isolation, so in the present study their predictive power was examined in the context of how they are used by admission committees in conjunction with other preadmission data, primarily uGPAs. Analyzing the MCAT in this context allowed examination of the incremental predictive validity of MCAT scores used in addition to uGPAs. In order to examine the predictive power of different combinations of preadmission variables, five predictor sets were used in the analyses:

- uGPAs (undergraduate science and nonscience GPAs) alone,
- MCAT scores (Verbal Reasoning, Biological Sciences, Physical Sciences, and Writing Sample) alone,
- MCAT scores in conjunction with uGPAs,
- uGPAs and undergraduate-institution selectivity, and
- MCAT scores, uGPAs, and undergraduate-institution selectivity.

The criterion variables were medical school year 1 and year 2 GPAs, which were initially analyzed separately, and subsequently combined into a cumulative GPA, which is reported here; year 3 GPA; USMLE Step 1, Step 2, and Step 3 scores; and incidence of academic difficulty and distinction. During the early 1990s, which is the time period for the cohorts whose data were used in this study, most medical schools' course requirements for years 1 and 2 were the same or very similar, and the year 3 requirement was usually the required clerkship rotations. So, even though the grading procedures differed across schools, there is some commonality in the subject matter the criterion grades are intended to represent.

Procedures

This study followed two cohorts from entrance to medical school through residency. Students from 14 medical schools' 1992 and 1993 entering classes provided data for predicting medical school grades and academic difficulty/distinction, while their peers from all of the U.S. medical schools were used to predict performance on USMLE Steps 1, 2, and 3. The 14 schools were selected so that as a group, the sample was representative geographically, racially, and ethnically of U.S. medical schools.*

The sample also represented public and private schools and schools with varied curricular approaches, such as traditional, systems-based, and problem-based curricula. The number of medical students in a school's cohort ranged from 65 to 148, with a median of 126 for the 1992 cohort and 107 for the 1993 cohort. Course grades for a total of 4,076 students in this "medical school sample" were collected from these schools. For most analyses, I report results for the combined 1992 and 1993 cohorts and refer to this aggregation as the "combined cohorts."

For the prediction of licensure examination (Step) scores, my colleagues (see acknowledgment) and I used the scores of all of the more than 31,000 students who entered the 125 U.S. allopathic medical schools in 1992 and 1993 and refer to it as the "national sample." Students progress through their medical education and the Step examinations at different paces. Some reach Step 3 within six years of sitting for the MCAT, others take much longer, and some leave medical school before finishing.[†]

As a result of the delays and attrition, a decreasing number of students reach each milestone in their medical education within a given time period, such as that encompassed by this study. The national sample contained Step 1 scores for 27,406 students, Step 2 scores for 26,752, and Step 3 scores for 25,170, as of 2002, when the final analyses were run.

Descriptive and regression analyses were used to determine the nature and strength of relationships among the variables of interest. Because each medical school had its own criteria for assigning grades, analyses on the medical

*Medical schools participating in the study were the University of Arizona College of Medicine; Wake Forest University Health Sciences (School of Medicine); University of California, Los Angeles, David Geffen School of Medicine; Creighton University School of Medicine; University of Texas Health Science Center at San Antonio; Howard University College of Medicine; University of Illinois at Chicago College of Medicine; University of Oklahoma College of Medicine; University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School; University of South Carolina School of Medicine; Southern Illinois University School of Medicine; Stony Brook University Health Sciences Center; University of Virginia School of Medicine; and Washington University School of Medicine.

[†]In their analysis of the national sample of 1992 cohort, Huff and Fang⁷ found that 4% of medical students quit or did not graduate on time and cited academic difficulty as the reason.

school sample were completed separately for each medical school. Results were summarized across schools by reporting the median and ranges of the correlation coefficients. The Step scores' national scaling allowed us to predict them based on the entire group of medical school students in the national sample.

Description of the variables

This section describes the preadmission variables used in combination as the predictor variables, described above, and the outcome measures used as criteria.

uGPA. The term "uGPA" indicates the average of all undergraduate course grades, not including postbaccalaureate and graduate courses. Each medical student's science GPA (grades in biology, chemistry, physics, and mathematics) and nonscience GPA (grades in all other subjects) were extracted from the student's application records stored in AAMC databases. During the medical school application process, these GPAs were verified against undergraduate transcripts and standardized to adjust for differences in course lengths (e.g., semesters versus quarters) and definitions of "science" courses. Science and nonscience GPAs were entered as separate variables, but in the same step of

the regression equation, as a block, and will be jointly referred as "uGPAs."

MCAT scores. All four MCAT scores were used: Verbal Reasoning (VR), Biological Sciences (BS), Physical Sciences (PS), and the Writing Sample (WS)—jointly referred to in this report as "MCATs." The three multiple-choice sections of the MCAT (VR, BS, and PS) and the Writing Sample are designed to assess (1) mastery of basic concepts in biology, chemistry, and physics; (2) facility with scientific problem solving, and critical thinking; and (3)

communication/writing skills. The range of possible scores for BS and PS was 1–15; for VR, 1–13; and for WS, 1–11; the WS range was obtained by assigning numeric values to the letter-score range of J–T. The four MCAT scores were entered into the regression equation as separate variables, but all at the same time, as a block. For each individual in the study, only the most recent set of MCATs at the time of application to medical school was included in the analyses. This reflects the most common use of MCAT scores by medical schools.

Undergraduate-institution selectivity. The Astin Index is the average combined SAT score for all individuals admitted to a

particular institution. It is commonly considered an indicator of the selectivity of an undergraduate institution and serves as a proxy for academic quality. The index reflects a characteristic of the institution, not the individual, so all individuals who graduated from a particular undergraduate institution have the same value for their selectivity index.

Medical school GPAs. Year 1, 2, and 3 GPAs were treated as weighted averages, which were created from information provided by the 14 participating medical schools. For each medical student, an end-of-year GPA was created by multiplying each course grade or each clerkship rating by the number of contact hours (or weeks) for that course, summing the weighted grades across courses, then dividing this sum by the total number of contact hours for that year. Students were not penalized for grade data that were missing; that is, the total number of contact hours was adjusted to reflect the number of courses for which the student had grades. The resulting averages were converted to a common 4.0 scale. When the individual-school analyses showed no consistent pattern of differences between the ability to predict year 1 and 2 GPAs, for the sake of brevity, a cumulative GPA was computed. The cumulative GPA is the average of the year 1 GPA and the year 2 GPA and is used in their place throughout this report. If students had any grades reported for the second year, they are included in the analysis. The exclusion of those who dropped out of medical school before beginning the second year results in a more conservative analysis, because those most-likely-poor performers are not available for the predictor variables to successfully identify.

USMLE Step 1, 2, and 3 scores. The USMLE is a three-step examination that assesses knowledge and skills that medical students acquire throughout their medical education. Medical students must pass all three USMLE examinations to obtain licensure in the United States. Step 1 is often taken after the first two years of medical school and focuses on the understanding and application of basic science areas that are relevant to medical education. Step 2, taken at the

List 1

Academic Tracking Codes, by Category, Used to Provide Information about Students' Progress throughout Medical School*

Category A: Distinction in Medical School

- General Academic Distinction in Course Work
- General Academic Distinction in Clerkship Setting
- Research Distinction
- Community Service Distinction
- Leadership Distinction
- AOA
- Other Distinction

Category B: Leave of Absence

- Academic Remediation
- Other Academic Reasons

Category C: Change in Expected Graduation Date

- Academic Difficulty in relation to Course, Clerkship, or Electives Requirements
- USMLE Examination Failure
- Institutional Comprehensive Evaluation Failure
- Other Academic Reasons

Category D: Withdrawal from Medical School

- Academic Difficulty
- Other Academic Reasons

Category E: Dismissal from Medical School

- Academic Difficulty in Relation to Course, Clerkship, or Electives Requirements
- USMLE Examination Failure
- Institutional Comprehensive Evaluation Failure
- Other Academic Reasons

*Eleven of the 14 medical schools that participated in this study provided academic tracking information about each of their students who matriculated in 1992 and 1993, based on the list of codes above.

Table 1

Validity Coefficients, Corrected (and Observed) for Range Restriction, for Predicting Medical School Grades and United States Medical Licensing Examination Step Scores for Students in the 1992 and 1993 Cohorts

Preadmission predictor set	GPA, medical school sample*		United States Medical Licensing Examination, national sample†		
	Cumulative‡	Year 3	Step 1	Step 2	Step 3
Undergraduate GPAs alone[§]					
Median	.54 (.40)	.36 (.26)	.49 (.39)	.44 (.33)	.42 (.29)
Range of school coefficients	.42–.76 (.06–.55)	.33–.56 (.18–.37)			
MCATs alone[§]					
Median	.59 (.44)	.46 (.32)	.70 (.61)	.60 (.49)	.62 (.49)
Range of school coefficients	.38–.78 (.20–.70)	.31–.54 (.18–.41)			
Undergraduate GPAs and MCATs^{§,¶}					
Median	.71 (.53)	.54 (.39)	.72 (.64)	.63 (.52)	.64 (.52)
Range of school coefficients	.53–.81 (.37–.73)	.42–.65 (.28–.49)			
Undergraduate GPAs and Select^{§,¶}					
Median	.60 (.45)	.42 (.29)	.57 (.47)	.51 (.36)	.48 (.31)
Range of school coefficients	.49–.79 (.31–.60)	.33–.62 (.22–.43)			
Undergraduate GPAs, Select, and MCATs^{§,¶}					
Median	.71 (.57)	.53 (.39)	.72 (.63)	.63 (.49)	.65 (.51)
Range of school coefficients	.55–.82 (.39–.73)	.43–.66 (.30–.50)			

* The *n* of students in 14 individual schools' cohorts ranged from 65 to 148, with a median of 126 for the 1992 cohort and 107 for the 1993 cohort. Data, for a total of 4,076 students in the combined cohorts, were collected from these schools.

† The *n*'s for the national sample = 27,406 for Step 1, 26,752 for Step 2, 25,170 for Step 3. See the text for the definition of the national sample.

‡ Cumulative GPA = GPAs from year 1 and year 2 of medical school.

§ For each pair of numbers in each column, the first number is the coefficient corrected for restriction in range. The second number (in parentheses) is the observed coefficient.

¶ "Select" means undergraduate institution selectivity index.

end of medical school, assesses clinical skills knowledge and whether the student can apply it under supervision. Step 3 is taken after a year of residency and assesses whether the graduates can independently apply their knowledge and skills. The three-digit USMLE scores range between 140 and 280, with a mean of first-time takers between 200 and 220 and a standard deviation of 20. The numeric scores are reported along with the pass/fail decision. Students' first attempt at each Step examination was used in these analyses.

Academic difficulty and distinction.

Academic success can be measured in several different ways. End-of-year GPAs and USMLE total scores are relative-standing criterion measures, where a student's success is measured in part by how well she or he is doing in comparison

to her or his peers. Another way to measure success in medical school is to track whether a student graduates on time, achieves academic honors, or encounters academic difficulty. The present study provided a unique opportunity to gather detailed information regarding students' progress throughout medical school. Eleven of the 14 participating medical schools provided academic tracking information about each student in the study, based on the codes shown in List 1.

Using these codes, three categories of students were identified: students who earned general academic distinction, students who earned distinction in clerkships, and students who encountered academic difficulty. The latter group included only those whose stated reason for leave of absence, change in graduation date, withdrawal, or dismissal was academic difficulty.

Results

Overview

Table 1 presents the results of multiple regression analyses using five combinations of preadmission variables, or predictor sets, to predict outcomes such as medical school GPAs and Step scores. The multiple correlation, or validity, coefficient is a measure of the strength of the relationship between the predictor set and the criterion measure and ranges in value from 0 (no relationship) to +1 (perfect relationship).

The observed relationship is an underestimate of what would have been found if the entire spectrum of applicants had been admitted to medical school. Since applicants with low MCATs or GPAs are less likely to be admitted, the range of values in the student population is generally reduced (i.e., restricted).

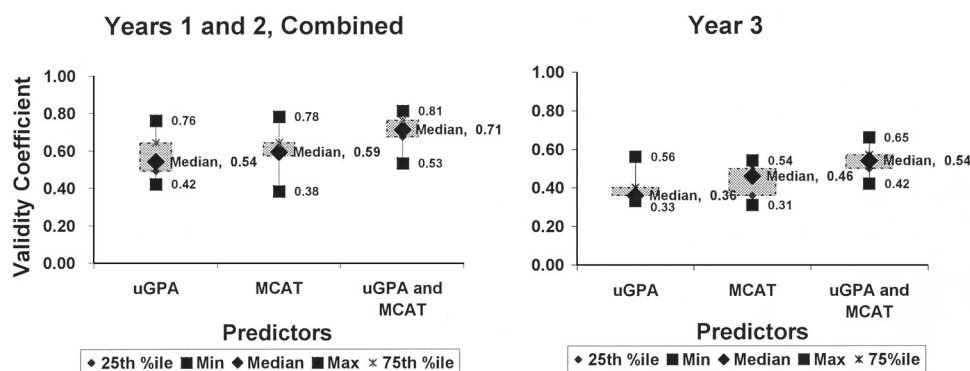


Figure 1 Validity coefficients obtained when preadmission data were used to predict medical school GPAs. This box-and-whiskers plot shows the range and density of the distribution of the validity coefficients of the 14 schools that supplied the data. The open box in the center of each distribution joins the 25th and 75th percentile points. The coefficients for half of all schools lie within this box. The figure demonstrates how the predictive value of all preadmission information varies among medical schools. (The term “uGPA” means “undergraduate GPAs.”)

After adjustment for restriction in range, described by Lord and Novick⁸ as being the ratio of the standard deviation of students' scores on the predictor variables to that of all applicants, the corrected validity coefficient estimates the strength of the relationship if there had been no selection. Coefficients of .40 or higher indicate a fairly strong relationship between the two sets of variables. Table 1 reports both observed validity coefficients and coefficients that have been corrected for restriction in range.

The first and second columns of Table 1 (under “GPA, medical school sample”) display the validity coefficients obtained when the different sets of preadmission variables were used to predict both cumulative GPAs and third-year clerkship grades. The columns contain the median and range of corrected coefficients obtained for the set of participating schools, with the observed values in parentheses. The third, fourth, and fifth columns similarly show the validity coefficients obtained when the preadmission variable sets were used to predict the three USMLE Step scores.

When the validity coefficient is squared (R^2), it indicates the percent of variance in the criterion measure explained by the predictor set. For example, the median validity coefficient of .59 for MCATs predicting cumulative GPA indicates that MCAT scores alone explain approximately 35% (0.59^2) of the variation in cumulative GPA.

Comparison of validity coefficients across predictor sets provides information about the added value of a particular set of variables. For example, comparing the validity coefficient for predicting

cumulative GPA from uGPAs alone (.54 for predictor set 1) to that obtained after the addition of MCAT scores (.71 for predictor set 3) shows an increase from 29% of the variance explained to 50%, revealing that the combination is a more powerful predictor of grades than is either set of information alone.

Corrected validity coefficients are displayed graphically in Figure 1. The plot on the left side of the figure shows the validity coefficients obtained when uGPA alone, MCATs alone, and uGPA and MCATs in combination were used to predict cumulative GPA; the right side reflects the prediction of year 3 GPA. The selectivity index is not included in this figure because it adds little to prediction after MCATs are included (compare the difference in Table 1's results for predictor sets 3 and 5).

The range of validity coefficients across the 14 medical schools is one of the more profound results of this study: the predictive value of all preadmission information varies among medical schools, implying that uGPAs and MCAT scores are much better predictors of performance at some medical schools than at others.

Cumulative GPA

Among the 14 schools, the highest corrected validity coefficient, using MCATs and uGPAs as predictors of cumulative GPA, is .81. For the school with that coefficient, almost two-thirds of the variance ($.81^2 = 66\%$) in cumulative GPA is explained. The lowest school's corrected validity coefficient is .53. With only 28% of the variance explained, this is less than half as much.

The median corrected validity coefficient for MCATs and uGPAs together predicting cumulative GPA is .71, which means that, on average, about 50% of the variance in cumulative GPA is explained by MCATs and uGPAs together ($.71^2 = .504$). By comparing the squared validity coefficients from Table 1, we can see a 21% gain by using these two preadmissions variables together, compared to using uGPAs alone, with that 21% representing the incremental validity from the addition of the MCAT.

The median incremental increase in the validity coefficients as a result of adding MCAT scores to uGPAs is .17 ($.71 - .54$). While seemingly modest, these increases almost double the proportion of variance explained by uGPAs alone (from $.54^2 = 29\%$ to $.71^2 = 50\%$). Adding uGPAs to previously entered MCAT scores yielded an incremental validity of only 15% (from 35% to 50%).

Year 3 GPA

Among the 14 schools, the highest validity coefficient for the prediction of year 3 GPA is .65. For that school, 42% of the variance is predictable using uGPAs and MCATs. The lowest validity coefficient using these same variables is .42, where only 18% of the variance in year 3 GPA is explained. The median increase in the validity coefficients as a result of adding MCAT scores to uGPAs is .18.

The median corrected validity coefficient using both MCATs and uGPAs as predictors of year 3 GPA is .54 with about 29% of the variance explained. The incremental gain from adding MCATs to uGPAs is more than half of that 29%

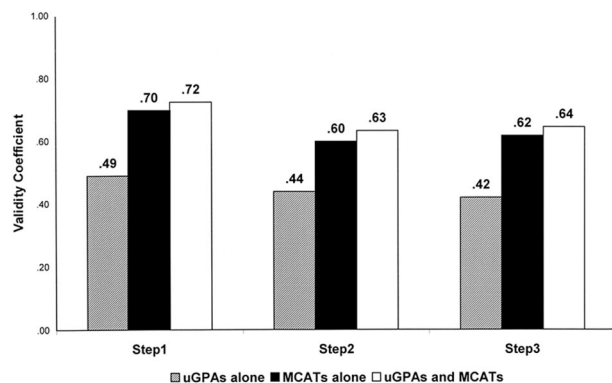


Figure 2 Prediction of United States Medical Licensing Examination (USMLE) Step scores. The bar graph depicts the validity coefficients for USMLE Step 1, 2, and 3 scores. The figure shows that the incremental increase in the percentage of variance explained in Step scores sometimes more than doubles when MCATs are used with undergraduate GPAs (uGPAs) as predictors, compared with when uGPAs are used alone.

(from 14% to 29%). MCAT scores alone predicted 21% of the variance. Although grades in the clinical third year are the least predictable of the criterion variables, using MCATs in addition to uGPAs more than doubles the variance explained. The biggest impact of using MCATs in addition to uGPAs to predict medical school grades is in the prediction of year 3 GPA.

Step scores

USMLE Step scores are reported on the same scale for all medical students, so Table 1 does not report the median and the range for the schools' Step validity coefficients; instead it reports the national observed and corrected validity coefficients for the Step examinations.

MCATs and uGPAs each contribute something unique to the prediction of medical school grades, and so the combination is more powerful than is the case with either predictor alone. In contrast, the contribution of uGPAs to USMLE Step scores is largely subsumed by that of MCAT scores, and so adds little to the predictive power of the combination. For example, for the prediction of Step 3 scores, the corrected validity coefficient using MCATs and uGPAs combined is .64 compared to .62 for MCATs alone and only .42 for uGPAs alone. MCAT scores alone are almost as accurate as predictors of USMLE scores as are the two together.

The increase in the validity coefficients as a result of adding MCAT scores to uGPAs for the prediction of USMLE Step scores are .23, .17, and .22 for the Steps 1, 2 and 3, respectively. Figure 2 shows that the

incremental increase in the percentage of variance explained in Step scores sometimes more than doubles when MCATs are used with uGPAs as predictors, compared with when uGPAs are used alone. Again, predictor sets that included institutional selectivity are not displayed because selectivity contributed little to the prediction of Step performance beyond MCATs and uGPAs.

Prediction of academic distinction and academic difficulty

The six plots in Figure 3 combine data across 11 schools and separate the MCAT content areas. For each of these plots, MCAT scores appear on the horizontal axis, and the percentage of students falling into each category (academic distinction or difficulty) is noted on the vertical. All of the Figure 3 plots show the *percentage* of students in each MCAT score category that experienced academic difficulty. For instance, in Figure 3f, though almost 11% of students with Verbal Reasoning scores below 4 experienced academic difficulty, the total number of students in this score category is only 66 nationwide (of whom seven experienced academic difficulty).

Plots 3a and 3b show the distribution of academic distinction and difficulty by BS and PS scores. In plots 3c and 3d, for the VR and WS scores, general academic distinction is separate from distinction during the clerkship, with the thought that these MCAT content areas might be more related to the abilities that are associated with clerkship distinction. The plot labeled 3e shows the combined cohorts' relationship of difficulty and distinction to average MCAT score.

A general pattern emerges across this series of figures; the proportion of students experiencing academic difficulty decreases as MCAT scores, in all of the content areas, increase. To a lesser extent, there is a pattern of increasing probability of distinction as MCAT scores increase. It should be noted, however, that incidents of distinction occur for students with very low MCAT scores, and incidents of difficulty occur for students with very high MCAT scores. Clearly, student characteristics unrelated to the knowledge, skills, and abilities measured by the MCAT are affecting the students experiencing difficulty or being awarded distinction.

The final plot, labeled 3f, uses data from all 14,275 students with complete data (from a total of 16,289 matriculants) who matriculated in 1992, not just those from the 11 schools. These data are from a study by Huff and Fang.⁷ They considered that a student was having academic difficulty if he or she withdrew due to academic reasons, was dismissed due to academic reasons, or delayed graduation due to academic reasons (other than MD/PhD programs, research leaves etc.).

Discussion

Medical school grades and licensure examination scores are different types of criterion variables. GPAs are school dependent, while the examination scores are not. Each school assigns grades according to its own criteria, where the "curve" references students from only that school. For instance, a performance that earns an "A" in one medical school might earn only a "B" in another. For that reason, all of the GPA prediction studies were conducted within each school. Medical school grades are not generally available in AAMC databases, so they had to be specifically collected and standardized, which is why the study was limited to a carefully selected, representative group of 14 schools.

In contrast, the licensure examination creates scores for all medical students on the same scale. This common scale allowed us to include students from all medical schools in a single analysis. As a way of determining whether the different levels of analysis contributed to the different results for the two criteria, my colleagues and I analyzed the licensure

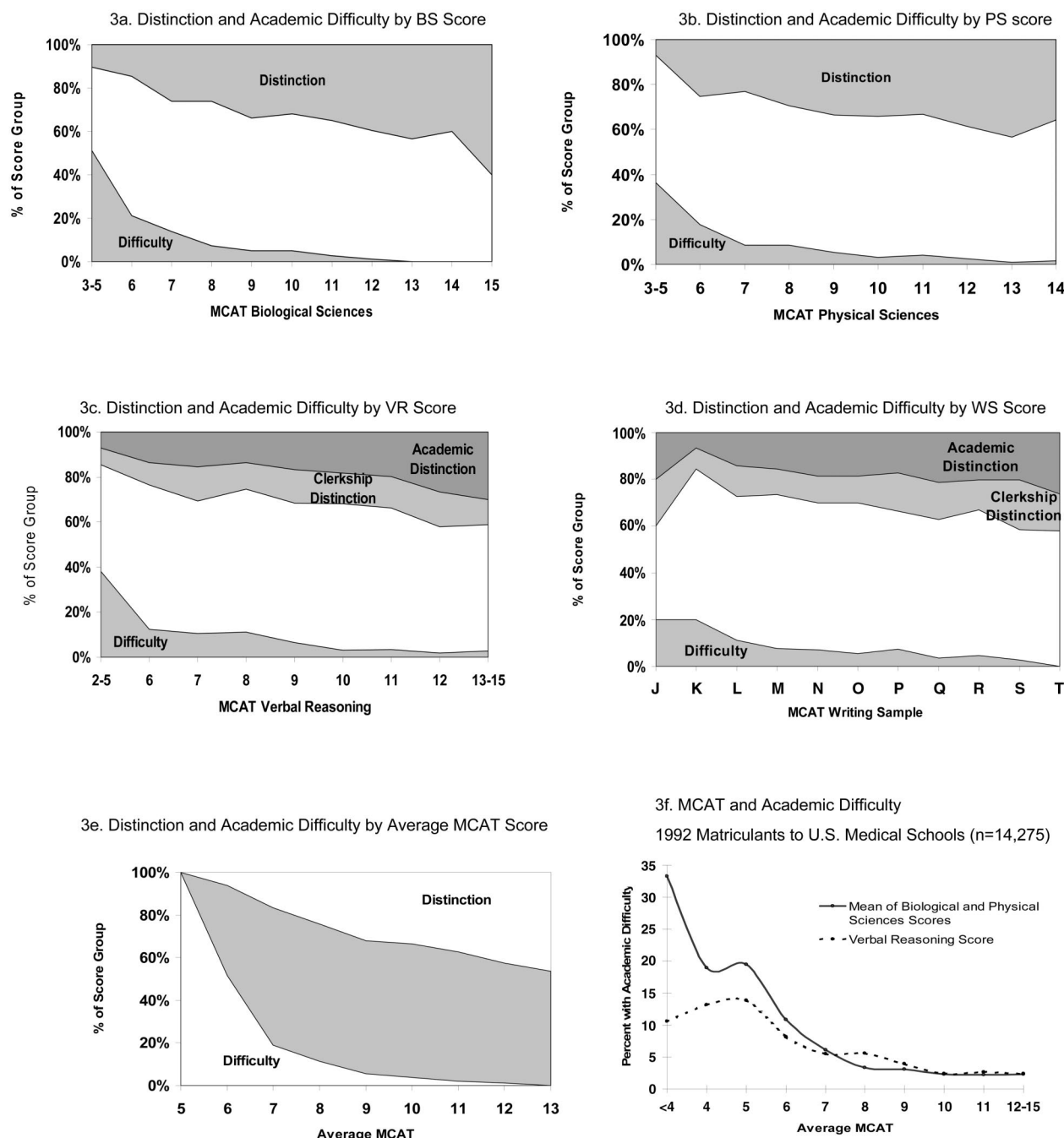


Figure 3 Students' distinction and academic difficulty compared to MCAT scores on the four test sections and to average MCAT scores. The data in all panels except 3f are from 4,076 students who were matriculants at 11 medical schools in 1992 and 1993; panel 3f uses data from all 14,275 U.S. students with complete data (from a total of 16,289 matriculants) who matriculated in 1992. This figure illustrates that the proportion of students experiencing academic difficulty decreases as MCAT scores, in all of the content areas, increase, although it should be noted that there are exceptions to this association. In panel 3d, Writing Sample scores are reported as letters, ranging from J to T.

examination data both ways—the validity coefficients were calculated individually for each of the PVR schools, as well as for the entire national population. The medians of the 14 schools corresponded closely to the national coefficients, which was seen as confirmation of the use the median validity coefficient for the schools as a proxy for the national relationship of MCAT scores to medical school grades.

The other relevant difference between the two types of criterion variables is that the USMLE licensure examinations are multiple-choice examinations, similar in format to that of three of the four sections of the MCAT. Any variation among persons that is unique to performance on multiple-choice questions would be present and consistent on both the MCAT and the

Step scores, increasing their correlation. While it might be tempting to discount the relationship between the MCAT and Step scores as an artifact, this would be reasonable only if the Step examinations were being used as a proxy for "medical school performance." In reality, however, performance on the Step examinations is necessary in its own right, and prediction of that performance, including whatever

variation is induced by methodological factors, is a valid and important role of the MCAT.

The prediction of licensure examination scores is also a role that the MCAT performs quite well. Even Step 3, typically taken six years after the MCAT, is predicted with greater accuracy than are first-year basic science GPAs. One reason is certainly the common element of being multiple-choice, high-stakes examinations. Another reason for the stronger relationship between MCAT scores and Step scores than between MCAT scores and GPAs is that no examination can measure all the elements that are involved in producing course grades.

Not only do the grading criteria vary among medical schools, but so also do the relationships among the knowledge, skills, and abilities a student brings to medical school and what they achieve. To a large extent for medical school GPAs and to a lesser extent for Step scores, the predictive value of all of the preadmission variables varies dramatically. For some schools in our study, the students who ranked highest on uGPAs and MCATs also achieved the highest grades and Step scores. For other schools, there was little relationship: those who entered medical school with weaker academic credentials were as likely to achieve high grades and Step scores as those who came in with stronger credentials. Discussion with these schools, usually public schools with a strong commitment to recruiting state residents, suggests that they use MCAT scores as diagnostic tools to predict which students will need extra assistance, and then they work hard with the at-risk students to enable them to be successful in medical school. They use MCATs to predict who is going to need help to succeed, not to predict who will succeed. This seems like an important and valid use of MCAT scores, the value of which will not be evident in studies of predictive validity. In fact, the more the MCAT is

successfully used this way, the less predictive validity it will show. The other explanation offered by the schools participating in this study was that the lower-scoring students whom they admit exhibit other exceptional qualities that make them more able to overcome their initial academic deficits.

Moulder and colleagues⁹ conducted research to determine what characteristics of medical schools are most related to the predictive power of undergraduate information. The school characteristics in that study—size, public versus private, region of the country, and selectivity—are the same ones that the present study attempted to balance among the 14 validity-study schools. The present study also included representatives of both traditional and problem-based learning curricula; the school characteristics study did not. Because one half of the MCAT focuses on science, it might be expected to better predict grades in courses that cover “purer” sciences, which tend to be in more traditional curricula. On the other hand, the MCAT’s strong emphasis on problem solving in novel situations, and the fact that half of the MCAT focuses on verbal skills suggests that it might also predict well in more self-directed learning environments.

The medical schools’ strong or weak relationships between preadmission variables and performance may be reflections of their schools’ missions. For some schools, the priority is to graduate National Institutes of Health researchers; for others the priority is to provide physicians for rural and underserved areas. The performances that they encourage and reward with good grades differ, as do the relationship of their students’ MCAT scores to their performance in medical school. Both missions are important and enrich the nation’s pool of physicians by their very difference. The importance of those differences in both missions and

regression results requires caution in the citing of a single number as representative of “the” relationship between MCAT and medical school performance.

The listed author wishes to make it clear that showing only one author misrepresents the long-term collaborative nature of this work. More than two dozen MCAT staff and medical school volunteers made this project happen. Most noteworthy of these are Judith Koenig, who was instrumental at both the beginning and ending of this almost 15-year journey, Karen Mitchell, who started it all, Kristen Huff and Andrew Wiley, who got and kept it going, and Patricia Etienne, who finished it.

References

- 1 Mitchell KJ, Haynes R, Koenig JA. Assessing the validity of the updated Medical College Admission Test. *Acad Med*. 1994;69:394–401.
- 2 Wiley A, Koenig JA. The validity of the Medical College Admission Test for predicting performance in the first two years of medical school. *Acad Med*. 1996;71(10 suppl):S83–85.
- 3 Koenig JA, Sireci SG, Wiley A. Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Acad Med*. 1998;73:1095–106.
- 4 Hojat M, Erdmann JB, Veloski JJ, et al. A Validity Study of the Writing Sample Section of the Medical College Admission Test. *Acad Med*. 2000;75(10 suppl):S25–27.
- 5 Bridgeman B, Pollack J, Burton N. Understanding What SAT Reasoning Test Scores Add to High School Grades: A Straight forward Approach. (Research Report No. 2004-4). New York: The College Board, 2004.
- 6 Kuncel N, Hezlett S, Ones D. A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychol Bull* 2001;127:162–81.
- 7 Huff KL, Fang D. When are the students most at risk off encountering academic difficulty? A study of the 1992 matriculants to U.S. medical schools. *Acad Med*. 1999;74:454–60.
- 8 Lord FM, Novick MR. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.
- 9 Moulder B, Algina J, Julian E. Examination of the Effects of Medical School Characteristics on the Prediction of USMLE Results. MCAT Monograph #11. Washington, DC: Association of American Medical Colleges, 2003.