An Introduction to the Medical College Admission Test (MCAT)

Ellen Julian

Association of american Medical Colleges

History

The Medical College Admission Test (MCAT) began in 1928 (although it was then known as the Moss Test, after its original developer). It was created to address the problem of high attrition rates in medical school. In the early 20th century, the medical schools in the United States had attrition rates as high as 50% (McGaghie, 2002). By 1946, the attrition rate had reduced to 7%. The MCAT has had different names and different formats over the past 75 years. Major changes to the structure and content of the MCAT were instituted in 1946, 1962, 1977 and 1991.

In some of those versions, the content included only academic subjects. In other times, it also included measures of general knowledge which assessed knowledge far beyond the realm of the sciences and beyond what is taught in academic courses. These changes reflect the tension between the goals of having physicians with a broad, liberal arts background, and the simultaneous desire to have them be superb scientists. The McGaghie (2002) article offers a thorough researching of the history of the MCAT.

Validation

After the introduction of the revised MCAT in 1991, the AAMC began investigating the MCATs ability to predict success in medical school and how that compares to, and adds to, the predictive power of undergraduate grade-point averages. When the longitudinal study was completed in 2000, the research showed that medical-school grades were best predicted by a combination of MCAT scores and undergraduate grades, but licensure examination scores were better predicted by

MCAT scores than by undergraduate grades, and the combination did little better than MCAT scores alone (see Appendix A for a summary).

Score Use

Medical school admission committees consider MCAT scores along with other information in evaluating qualifications for medical school. The emphasis that the committees place on MCAT scores varies from school to school. Other sources of information used by admission committees include:

- undergraduate grade-point averages;
- breadth and difficulty of undergraduate coursework;
- letters of evaluation from undergraduate advisors or others;
- personal comments on American Medical College Application Service (AMCAS) and / or institutional application forms;
- involvement in extracurricular activities such as student government and community service;
- involvement in and quality of health-related work and research;
- participation in other activities demonstrating motivation, responsibility, maturity, integrity, resourcefulness, tolerance, perseverance, dedication to service, or other relevant characteristics;
- medical school interview results; and at some institutions,
- state or county of legal residence

A survey of medical school admission officers indicated that the MCAT serves three main functions for the committees:

- identification of applicants who are likely to succeed in medical school and those likely to experience academic difficulty;
- assessment of applicants specific strengths and weaknesses;
- interpretation of applicants transcripts and letters of evaluation.

In practice, MCAT scores are compared with candidates college records. Significant discrepancies between the MCAT score on one of the science sections and grades earned in courses in that science area are noted and explored. A high science score may compensate for lesser science grade data. In the case of the Writing Sample, high-quality essays may help an admission committee feel more comfortable with any ambiguous information about communication / writing ability obtained through the application or interview.

MCAT scores are given greater attention

in evaluation of the academic records of candidates from colleges that are unfamiliar to the admission committee. In such cases, comparing MCAT scores with grades provides an estimate of the candidates academic accomplishments in relation to those of candidates from colleges more familiar to the committee. Admissions offices also are provided with a confidential document (on CD) containing the four-year rolling-average of average MCAT scores for every undergraduate institution that had at least 10 students take the MCAT during the four years.

The range of acceptable MCAT scores varies among schools (see Figure 1). The survey mentioned above asked admission officers to designate the lowest MCAT scores deemed acceptable of their institutions. At some schools scores of 4 were considered acceptable; at others only scores of 10 or more were considered acceptable. The average response for the lowest acceptable score was 7. The wide range in medical schools views of what are considered acceptable MCAT scores indicates considerable variation in selection criteria.

Content

Since 1991, the content of the MCAT has been approximately one-half science and one-half humanities. The Biological Sciences (BS) test section includes approximately 70% biology items (questions) and 30% organic chemistry. The Physical Sciences (PS) test section is approximately 52% general chemistry and 48% physics. The 1991 content blueprint for the MCAT was based on the results of the 1988 content review.

Undergraduate faculty in the sciences and humanities initially constructs MCAT questions. They select text passages from the published literature in their field, and construct 10 to 12 questions using the information contained in the passage to assess problem solving ability in areas specified by the content blueprint. The faculty are paid for each passage-question set produced. They work individually at their home or office, communicating with the contractor by mail. Work is underway to develop an Internet-based, secure system to assist the authors.

Content-specialist editors at the test-development contractor refine the questions submitted by the authors. Although the item authors have been well trained, some work to ensure that MCAT requirements are met is usually required. The passage-question sets are then submitted to a panel

of specially-trained bias reviewers, who look for evidence of content that might be offensive or stereotypical, or answers that might be obtained through external experiences, unrelated to preparedness for medical study, that is differentially available to groups of examinees. After each administration, statistical differential item functioning (DIF) analyses are conducted to determine if the items are functioning in the same way for all groups that are sufficiently large to support the statistics (usually male/female, black/white).

Content Review

A similar review was completed in 2001, but with the goal of making incremental, rather than fundamental, changes in MCAT content. The goal of the 2001 content review was to ensure that the science content is complete and current. A constituent committee, the Program Review Advisory Committee, determined that the presence and structure of the Verbal Reasoning and Writing Sample test sections were not to be changed. There was a discussion of combining the two science sections in order to put less emphasis on the sciences when a total score is computed. This change was deferred to a later date.

Both medical schools and undergraduate science faculty were involved in the content review process. A list of all topic areas covered in the major textbooks was compiled and developed into two surveys.

The Medical School Survey of Science Content (MSS, see Appendix B) was designed to identify science topics within the existing test specifications that were no longer important to success in medical school, as well as topics that were not represented in the current specifications, but which may be important to success in medical school. The MSS was sent to the Deans of all 144 U.S. and Canadian medical schools. The medical schools were asked to give the survey to two basic science faculty, two clinical science faculty, two residents and two students. The discipline of the basic science faculty was not specified. Within the sample, all disciplines were represented.

The MSS showed the list of all topic areas at two levels, the topic and subtopic levels. For each topic, respondents rated, on a five-point scale, how important they thought it was for entering medical students to have knowledge of that topic. They indicated any subtopics that were not important to know before arriving at medical school.

A random sample of 166 U.S. under-

graduate institutions participated in the Survey of Undergraduate Science Content (see Appendix C). This survey was designed to determine whether the science topics identified as being important to success were sufficiently covered in introductory science courses at the undergraduate level. Respondents rated the degree to which the topics are taught in undergraduate introductory courses. Topics with high ratings on the MSS but with low undergraduate coverage were not considered for inclusion in the new specifications.

Results were compared with those from the similar study conducted in 1988. Findings indicate that, as prerequisites for medical school, biology has increased in importance while the importance of chemistry, particularly organic chemistry, has declined in the past twelve years. As a consequence, three questions on DNA and genetics will be added to the Biological Sciences section, replacing three organic chemistry questions. This will change the balance within BS to 75% biology and 25% organic chemistry.

Test Format

In the science sections, the format is generally a passage (approximately 250 words) to be read, followed by a number of multiple-choice, select-the-best-answer, items (questions) assessing facility at problem solving using the information in the passage. In both science sections, there are several stand alone multiple-choice questions, in addition to the passages.

The Verbal Reasoning (VR) test section includes long (500-600 word) reading passages from the humanities, social and natural sciences, with between five and ten questions following. These questions are intended to go beyond reading comprehension (do you remember what you read?) to the assessment of the ability to understand, evaluate and apply information and arguments presented in prose texts. For example, a question might present additional information on the topic, and ask if this information would support or undermine the authors position.

The Writing Sample (WS) is two 30-minute essays on non-medical topics, where the scoring focuses on the examinees ability to craft a coherent argument, support it with examples, acknowledge the other side of the argument, and resolve the two positions. The essays are scored holistically, with regard to grammar and spelling only to the extent that they enhance or detract from the quality of the

argument.

Scoring

The multiple-choice section (BS, PS and VR) scores are reported on a scale that ranges from 1 to 15, with an average of approximately 8. The Writing Sample score is reported on an alpha scale that ranges from J to T, with J being the lowest score and T being the highest. The Writing Sample is reported as letters to make it difficult for users of the scores to average it with the other three scores. Its correlation with them is so low (see Table 1) that to combine it with the others results in the loss of information.

Table 1. Correlations among MCAT test sections, August 2002 (N=31,820)

Section	PS	VR	WS	Total (BS+PS+VR)
BS	0.79	0.63	0.37	0.92
PS		0.56	0.36	0.89
VR			0.38	0.83
WS				0.42

Historically, users were discouraged from combining scores from the other three sections. However, they often did so. Admissions officers, students and the literature (even our own research) frequently refer to a Total MCAT score. Beginning in 2003, the MCAT program will report a Total Score, the concatenation of the sum of the three multiple-choice sections and the Writing Sample, e.g., 45T, in addition to the four individual scores.

From 1992-2002, the highest score in Verbal Reasoning (VR) was reported as 13-15. In 1992, there were insufficient difficult questions in VR to enable distinctions among the highest levels of performance, and so 13-15 was the highest VR score reported. In succeeding years, a sufficient number of difficult questions were included, but this reporting convention was continued for two reasons. First, examinees from 1992 might be at a disadvantage compared to those from later years if they were unable to earn as high a score. Second, the equating of score scales across forms and years resulted in the 1992 score scale perpetuating itself across time. Beginning in 2003, scores of 13, 14 and 15 will be reported for VR.

The averages, standard deviations, and reliabilities of the test sections are shown in Table 2. These data are for the August 2002 administration. The Total score shown in the Table was not actually reported to the examinees, but represents what might be expected in future years.

Table 2. Descriptive statistics for MCAT test sections, August 2002 (N=31,820)

	Mean	SD	Reliability
BS	8.18	0.63	0.37
PS	7.88	0.56	0.36
VR	7.66		0.38
WS	6.20('O')		
Total	23.71	6.209	0.94

For the Writing Sample, the median alpha score (O) is usually reported, rather than the mean of the alpha scores numeric equivalents. However, for research purposes, the numeric equivalents are used. The reliability of the Writing Sample cannot be fully reflected in a single number. The correlation of the scores assigned by the two raters reading each essay is usually in the mid-.80s. Scores on an examinees two essays typically correlate in the high-.60s. A generalizability theory analysis shows that the main source of variance in examinees scores is the essay topic. Reducing the WS to one essay would result in a significant loss of information.

Test Equating

Each section of the MCAT is equated back to the scale of the first administration of the current test configuration, April 1991. This enables the comparison of scores earned in different years, and on different sets of questions (forms) within the same year. At each administration, multiple test forms are administered. Some are previously administered forms; others are new, composed of previously field-tested questions. Each forms has multiple possible sets of field test questions. Field test questions are newly written questions whose performance is being evaluated.

The forms are randomly assigned to examinees. This allows the comparison of the current groups performance to that of the previous group that took each of the previously used forms. Using equi-percentile equating, the table for converting each forms raw scores (number of correct answers) to scale scores (on the 1 to 15 scale) is adjusted based on this information. Equi-percentile is not the newest or most flexible type of equating, but it is sturdy, accurate and the current system has many redundancies and checks built into it.

Research is underway evaluating the use of item response theory (IRT) models on MCAT data. IRT would allow the reuse of equating questions out of the context of the forms in which they were originally administered. Not only does this allow more flexibility in constructing new forms, but it also enables computer adaptive testing (CAT). The biggest difficulty in implementing IRT

on the MCAT is finding the best method for dealing with the local item dependence resulting from the pssage-based item sets. IRT models that incorporate the dependence within a set of questions (e.g., partial credit models) do exist, but are more experimental than simpler models. Research shows that VR has the highest level of statistical dependence among questions.

Future Changes

The AAMC is making a number of minor changes in the MCAT in 2003. The announcement of these changes is attached as Appendix D. Some of these have already been mentioned. Another limits an applicants ability to withhold scores from the medical schools.

In August 2002, a form of the MCAT was administered via computer to 31 examinees in London. In 2003, all international MCATs will be administered on the computer. This opportunity will provide valuable information about the process of computerizing the MCAT. While it may provide interesting information about the psychometric performance of the MCAT on computer, that is not its intent. Intact, previously administered forms will be used, so the scoring scales developed on the national group can be applied.

Research into the more effective use of computers is underway. Computerized adaptive testing (CAT) offers the potential for significant shortening of the test, thereby reducing the cost of seat time at the computer-based testing center.

Two other motivations for shortening the test exist. One is the result of a nation-wide debate about the flagging with an asterisk (*) of scores achieved under non-standard conditions (usually extra time) by students who have petitioned for such accommodations because of a disability (usually a learning disability). One possible outcome of this debate is the provision to all examinees of more time for answering each question. If the MCAT were to do this and stay a one-day test, the number of questions would need to be reduced.

Another reason for shortening the test is the desire to introduce additional measures, while keeping the MCAT at a single test day. Admissions committees have requested assistance with the evaluation of personal characteristics, such as empathy, compassion, integrity, altruism and communication skills. Efforts are underway for the building of a test of Listening Skills.

Services to Examinees and Schools

MCAT registration will be only online, beginning in 2003. ACT in Iowa City, Iowa, who is the main vendor for the MCAT, runs the online registration system. Information about the MCAT is available primarily online (see MCAT Essentials, Appendix E). A simple brochure is mailed to undergraduate advisors, libraries, etc.

Undergraduate advisors are active constituents of the MCAT. They are often the main source of information for medical school applicants. While they do not have an official voice in AAMC policy making, their needs and opinions are given much weight. They serve as advocates for the students.

Each undergraduate institution may designate a single Chief Health Professions Advisor (CHPA) for the AAMC. This database currently includes approximately 1400 individuals. When students register for the MCAT, they tell us if they want their scores to be released to their advisor.

The AAMC has developed an Advisor Information Service, which has just migrated to the Internet. Its original purpose was to provide up-to-date information about students applications to medical school (where they have applied, where they have been accepted). It will be expanded this year to include the list of their students MCAT scores. Before this, the list was mailed to CHPAs after each MCAT administration.

The AAMC offers a centralized application system (AMCAS) that is used by almost all of the U.S. medical schools. Students complete their application online and tell AMCAS which medical schools should receive their application. MCAT scores are automatically included in AMCAS applications.

If applicants wish their scores to be sent to other schools or other individuals, they make a request through the MCAT Testing History (THx) Web. They will also be able to view their scores at the THx site as soon as the scores are available to the AAMC. This might be as much as two weeks before they receive their scores in the mail.

Most U.S. medical schools (116 of 125) will receive MCAT scores through AMCAS. The others will pick them up from a secure ftp site. Applicants also may print out an unofficial score report to send to other places. The report will contain a unique code that may be used to verify the contents on a secure Web site. Generally, the AAMC

is moving towards providing more of its services to examinees, advisors and admissions officers via the Internet, in a more automated manner and at lower cost.

References

Julian, E., & Lockwood, J. (2000). Predictive Validity of the Medical College Admission Test. *Contemporary Issues in Medical Education*, 3, 2.

McGaghie, William C. (2002) Assessing readiness for medical education: Evolution of the Medical College Admission Test. Journal of the American Medical Association, 288 (9), 1085-1090.