Title
Building a better MCAT: the design, redesign, and use of the medical college admissions test, 1971-1980

Author
Lening, Christopher

Publication Date
2006

Peer reviewed|Thesis/dissertation

Building a Better MCAT: The Design, Redesign, and Use of the Medical College Admissions Test, 1971-1980

by

Christopher Lening

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF ARTS

in

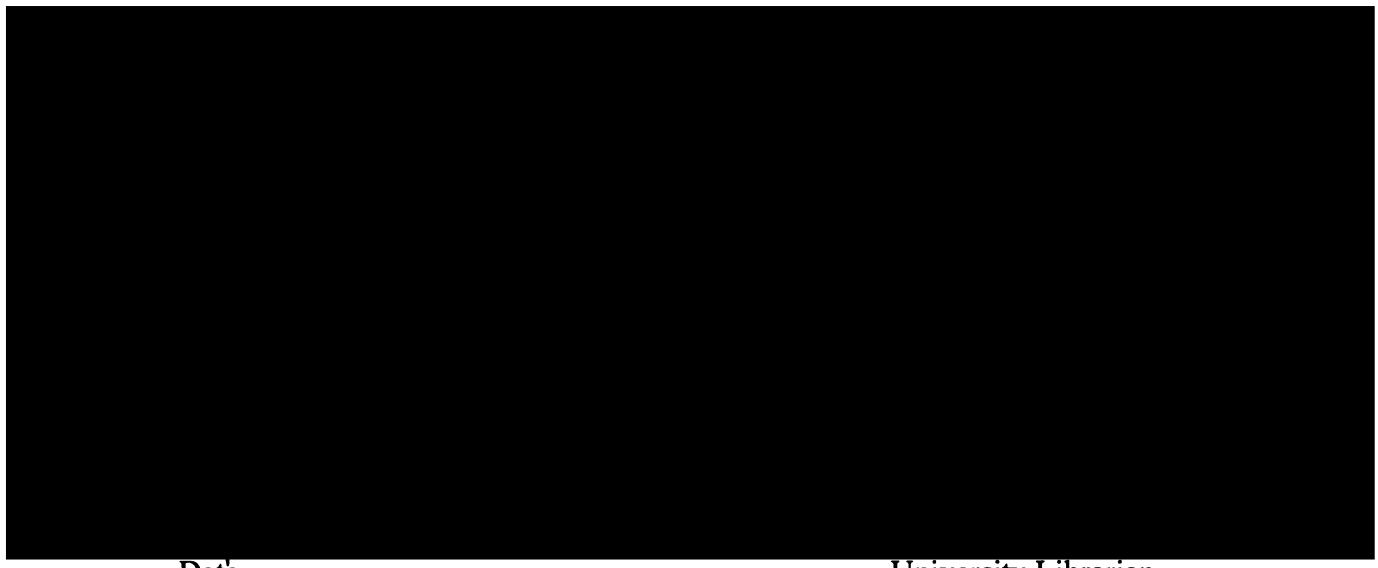
History of Health Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO



Date University Librarian

Degree Conferred:.....

ACKNOWLEDGEMENTS

I would like to thank the Department of Anthropology, History, and Social Medicine at UCSF for their support through this process, especially Akhil Mehra, Niki Nibbe, and Elizabeth Watkins for their suggestions during the initial stages of this process. Special thanks to Dorothy Porter, Brian Dolan, and Justin Suran for their instruction over the past year, and to Clair Dunne for being an invaluable resource.

Outside of UCSF, I would like to thank Naomi Rogers for proposing the idea, and Alexander LaFortune for helping me better understand the MCAT itself. My time in Washington was aided by the work of Kimberly Firth, Marian Taliaferro, and the rest of the library staff at the AAMC, and by the exceeding hospitality of Carey Lening and Katharine Tomlinson.

I thank my parents for putting up with me.

Building a Better MCAT: The Design, Redesign, and Use of the Medical College Admission Test, 1971-1980

Though many recent works examine the state of admissions in American education, undergraduate admissions have received almost all of the attention, leaving issues in post-baccalaureate admissions unexplored[1]. Medical schools present an intriguing case within this latter category. Thousands of high-achieving students apply for no more than a few hundred spots per school, each bearing a near guarantee of access to the profession. Unlike most other professional schools (most notably law schools), medical schools boast nearly negligible rates of student failure, dismissal, or other attrition; nearly all entering medical students will complete their doctorate and match with an internship program[2]. Thus, the admissions process represents the major point of discrimination between those judged capable of entering the profession and those judged incapable. Since the first major surges in medical school applications began in the late 1960s, however, there have been many more qualified applicants than available places in medical school. This conundrum forced those involved in medical school admissions to reexamine and adjust the methods used to select which students would become physicians.

---

[1] Recent notable examples include Nicholas Lemann, *The Big Test: The Secret History of the American Meritocracy* (New York: Farrar, Strauss, and Giroux, 1999), Jacques Steinberg, *The Gatekeepers: Inside the Admissions Process of a Premier College* (New York: Viking Adult, 2002), Jerome Karabel, *The Chosen: The Hidden History of Exclusion at Harvard, Yale, and Princeton* (Boston: Houghton Mifflin, 2005), Malcolm Gladwell, "Getting In: The Social Logic of Ivy League Admissions" *The New Yorker* 10 Oct. 2005, p. 80-86, and Alexandra Robbins, *The Overachievers: The Secret Lives of Driven Kids* (New York: Hyperion, 2006). The latter extends a look into preschool admissions, as well.

[2] Attrition rates can be found in the annual "Education Issue" of the *Journal of the American Medical Association*, published at varying points in the year. The most recently published attrition rate is found in Barbara Barzansky and Sylvia I. Etzel, "Undergraduate Medical Education" *JAMA* 290, no. 9 (3 Sep. 2003) p. 1190-6; out of 66,219 students enrolled in 2001-02, 673 left their medical schools. 133 of these students transferred elsewhere; the remaining 540 dropped out entirely, for reasons ranging from academic difficulty to serious illness.

One of the primary admissions tools, the Medical College Admission Test, lies at the center of this story. The MCAT is one of anywhere between three and seven factors listed as essential to medical school admissions, though it is only rarely considered the single most important determinant[3]. Sponsored by the Association of American Medical Colleges (AAMC) and currently administered by the American College Testing Program (ACT), the MCAT performs a not entirely dissimilar role to that of the other standardized admissions tests[4]. The MCAT represents the single most standardized part of admissions; relying on carefully selected, carefully evaluated questions offered to students only two weekends a year, the MCAT offers a unique comparison between applicants from different schools with different experiences and backgrounds[5]. Exactly what it compares, however, has been less readily apparent, leading to persistent controversy and the occasional complete overhaul. The most radical of these revisions occurred in 1977, as new pressures from the applicant pool and from medicine itself forced the AAMC to rethink the objective of the test. In the process, the structure and content of the test changed around the MCAT's new intentions, dropping broad evaluations in favor of a

[3] The MCAT, GPA, and interview are the only three items that appear consistently in every mass-marketed guide to medical school admissions; other factors cited include letters of recommendation, extracurricular activities and work experience, the primary AMCAS application, and secondary applications requested by individual schools. See, e.g., *Getting into Medical School: A Strategic Approach* ed. Larissa Shmailo (New York: Kaplan Publishing, 2003). Malaika Stoll, *The Princeton Review Best 162 Medical Schools* 2006 edition (New York, Random House, 2005), John Smart, Stephen L. Nelson, and Julie Doherty, *The Princeton Review: Planning a Life in Medicine: Discover If a Medical Career is Right For You and Learn How to Make It Happen* (New York, Random House, 2005). For comparative weight assigned to these elements by medical schools, see Karen J. Mitchell, "Use of MCAT Data in Selecting Students for Admission to Medical School" *JME* 62, no. 11 (Nov. 1987) p. 871-9.

[4] An internally produced history of the AAMC itself can be found in Mark D. Bowles and Virginia P. Dawson, *With One Voice: The Association of American Medical Colleges, 1876-2002* (Washington, D.C.: AAMC, 2003).

[5] The MCAT will move to a computer-based format in 2007, and allow administration on twenty-two occasions throughout the year. The paper-and-pencil MCAT was traditionally offered once in April, and once in August.

narrower band of material judged most relevant to the medicine's educational and practical needs.

These changes reflect the variety of uses that various groups had for the test. The AAMC sponsored and controlled how the test worked, and frequently attempted to use it as a way to influence the admissions process from the top down. The uniform, standardized nature of the test allowed it to be manipulated more easily than undergraduate grades, admissions interviews, or any other piece of the admissions puzzle. Regardless of the AAMC's intentions for the test, the ultimate meaning of it came from individual medical school admissions committees, who decided how, and in what ways, the MCAT could or could not be used in making their decisions. Frequently, the difficulties of assigning a limited number of places to an enormous number of qualified students led admissions committees to use the MCAT in unintended, ineffective ways, or curbed attempts to take the test in a different direction. Examinees also played a major role; their changing numbers and ranks created the problem that led to the introduction of a New MCAT, and their obsessive premedical preparations would produce much of the impetus for that test's further revision. Every group approached the MCAT differently, and one group's intentions for the test often directly conflicted with another's. There emerged a number of ways to the different MCATs given during this period, connecting groups to each other, and offering a model for greater trends in testing, medical education, and American medicine itself. As the simplest part of admissions to control, the MCAT has served as the battleground for larger conflicts over who physicians ought to be, and what was truly necessary for an aspiring doctor to know.

PREDICTING FAILURE

The introduction of a standardized test for medical school was an attempt to solve a particularly troubling problem in early twentieth century medical education. Through the 1920s, most medical schools lost at least twenty percent of their entering freshmen to failure, dismissal, or another form of terminal academic difficulty, and a few schools would lose as many as half of the students they accepted[6]. Most students left (or were forced to leave) by the end of the first year, and most of the remaining attrition occurred by the end of the second year; the expenses of facilities, supplies, and salaries in these years were being wasted on one in five students. Medical schools had no reliable way to predict which students were likely to end up in this bottom fifth.

Attempting to remedy the situation, the AAMC introduced the SAT for Medical Students in 1928[7]. The test came into use before anyone knew anything about its efficacy at predicting failure, drawing immediate controversy; the AAMC's Executive Council considered temporarily discontinuing the test in the early 1930, until more results were available.[8] Introduced only two years after the initial Scholastic Aptitude Test, the SAT-MS came with a fair risk it might be completely ineffective at its intended purpose. What this proto-MCAT lacked in demonstrated results, it made up for in scope; with several memory sections, specific subtests of scientific vocabulary and definition, and a subtest of "premedical information", the SAT-MS resembled neither the modern SAT nor the

---

[6] The original motives behind the MCAT are discussed in a brief historical overview by William C. McGaghie, "Assessing Readiness for Medical Education: Evolution of the Medical College Admission Test" *JAMA* 288, no. 9 (4 Sep. 2002), p. 1085-90. See also Charles B. Womer, "Past and Prologue" *JME* 56, no. 3 (Mar. 1981), p. 217-23.

[7] McGaghie, p. 1085.

[8] Womer, p. 219. The piece also recalls one skeptic who suggested the growth of tests and forms "would result in a great shortage of wood pulp."

MCAT that replaced it[9]. Most of these measurements seem to have failed, or at least proved less effective than those featured in the first "Medical College Admission Test" offered in 1946. Developed by the forerunner of the Educational Testing Service, the first version of the MCAT included four subtests: aptitude-based Verbal and Quantitative sections, resembling those seen on the SAT and other ETS creations, and two knowledge-based sections evaluating examinees in Science and "Understanding Modern Society."[10] This latter subtest aimed to promote a broad, liberal education to would-be medical students, but proved easy prey to accusations of bias against those on the margins of modern society. In 1962, a new iteration of the MCAT replaced "Understanding Modern Society" with the slightly broader, though equally controversial "General Information" subtest, evaluating knowledge of literature, history, religion, and more[11].

These three iterations of the test retained the common goal of predicting potential for academic difficulty, and seem to have achieved this goal adequately. Numerous studies demonstrated a greater tendency for below-average MCAT scorers to fail courses or be dismissed from medical schools, helping overall attrition rates fall under ten percent for the first time in the 1960s[12]. The MCAT's early identification of potentially problematic students allowed medical schools to opt instead for the by-now-demonstrated reliable performance of above-average examinees. In the decades after World War II, the

[9] McGaghie, p. 1087. The AAMC Archives do not include a copy of the SAT-MS, or any collection of exam-used questions prior to 1984. For the development of the SAT, see Lemann, p. 3-234. The SAT-MS used a unique scale, but the second and third iterations (the first to bear the MCAT name) borrowed and slightly modified the SAT 200-800 point scale, operating on ten-point intervals between 205 and 795.

[10] McGaghie, p. 1087, Lemann p. 70.

[11] McGahie, p. 1088. This version also shifted administrative responsibilities from ETS to ACT. The earliest extant study guides (offering non-AAMC-created practice tests) can be found for this version. See, e.g., William Gladstone, et al, *Handbook for the Medical College Admission Test* (New York, Arco Publishing Company, 1974).

[12] See James B. Erdmann, et al, "The Medical College Admissions Test: Past, Present, and Future" *JME* 46, no. 11 (Nov. 1971), p. 937-46, Joseph D. Matarazzo and Steven G. Goldstein, "The Intellectual Caliber of Medical Students" *JME* 47, no. 2 (Feb. 1972), p. 102-111, and Erdmann, "Separating Wheat From Chaff: Revision of the MCAT" *JME* 47, no. 9 (Sep. 1972), p. 747.

ratio of applicants to available places stayed at roughly two to one, and the MCAT proved a helpful guide in determining which half of the applicants was likeliest to graduate from medical school[13]. Since most medical school attrition occurred in the first two years, dominated by the basic science curriculum, the test needed only to predict preclinical performance, and thus centered on scientific knowledge rather than critical thinking[14]. The AAMC needed no further, higher-level delineation between those already deemed capable. As a result, the actual content of the questions was of secondary importance so long as graduates answered more of them correctly than did dropouts. In 1971, James B. Erdmann, director of Educational Research at the AAMC, admitted, "If the length of the big toe was found to be a better predictor than a highly relevant question in organic chemistry, preference would be given to toe length."[15] With no obligatory connection between the test and medical school curricula, the MCAT was useless as a diagnostic tool, and incapable of separating mediocre medical graduates from the top of the class. At a certain point above the mean, all scorers had roughly similar chances of graduation, since the test had been designed to separate a 655 from a 355, instead of from a 625 or even a 555. This system worked so long as all acceptably high scorers with adequate qualifications elsewhere had a place in medical school waiting for them.

---

[13] Application activity charts included complete data on the prior ten years of data only, but can be pieced together by combining multiple *JAMA* Education Issues, e.g. "Undergraduate Medical Education" *JAMA* 234, no. 13 (Dec. 29, 1975), p. 1336; Anne E. Crowley, et al, "Undergraduate Medical Education" *JAMA* 254, no. 12 (Sep. 27, 1985), and Barbara Barzansky, et al, "Educational Programs in U.S. Medical Schools, 1994-95" *JAMA* 274, no. 9 (Sep. 6, 1995), which combine to cover the period discussed in this paper.

[14] Erdmann, "Separating Wheat From Chaff", p. 747.

[15] Ibid. The practice test in Gladstone et al's review book may not provide an entirely accurate representation of the actual test, but it is the best of what was available. There are no questions on big toe size, and many of the questions on the science portion are not too far removed from the present content. However, the science subtest does include a question on the process delivering oxygen to the cells of a hydra (p. 22), and another on the metals that alloy to form brass (p. 24). Neither of these questions resembles anything that would be on a contemporary MCAT.

PREDICTING SUCCESS

In 1967, the number of applicants to American medical schools crossed twenty thousand for the first time; the number surged past thirty thousand four years later, and passed forty thousand only a year after that, finally cresting with 42,624 applicants in 1972[16]. Numerous theories attempted to explain this sudden surge; some medical educators boasted of the growing prestige of a medical career, while more skeptical observers noted that medical students could defer from the Vietnam draft[17]. The most straightforward explanation, though, was a sheer increase in the number of eligible students wrought by the growing number of college graduates; these numbers expanded further via a growing rate of female applicants and a concerted effort to increase minority medical school enrollments[18]. Although twenty-five new medical schools opened between 1966 and 1974, and several existing schools expanded their capacities, the number of applicants grew faster than the educational infrastructure would permit, forcing would-be medical students to reach a higher standard for consideration. Applicants with below-average grades or test scores were all but eliminated from the pool, and increasingly, so were students with merely adequate marks. The applicant surge drove medical schools to begin making the never-intended distinction between 655 and 555, and even 655 and 625, quickly exposing the MCAT's limitations.

---

[16] See the authorless "Undergraduate Medical Education" article in the 1975 Education Issue of *JAMA* from note 13, p. 1337. Medical schools take applications in one year for admission the next, so applicants filed in 1972 for a place in the 1973-74 freshman class.

[17] See, e.g., Jeremiah A. Barondess and Robert J. Glaser, "Attitudes toward the Medical Career: Findings from the Alpha Omega Alpha Survey of College and University Undergraduates" *Academic Medicine* [Hereafter *AM*] 68, no. 5 (May 1993), p. 323, and "Final Report of the AAMC National Task Force With Recommendations for the Medical College Admissions Assessment Program Study", in the AAMC Archives "Medical College Admissions Assessment Program: Box 1" Box 501982511, Folder 2.

[18] The number of 22-year-olds in a given year had been used as a rough estimate for the rise or fall of medical school applicants. The first baby boomers, born in 1946, turned 22 just after this increase began. See J.R. Schofield, "The Stork, Admission to Medical School, Going to a Foreign School, and Other Hazards" *JME* 48, no. 7 (Jul. 1973), p. 693-5.

The surge in applicants brought a growing number of statistically impressive candidates under consideration. The average GPA and MCAT scores of entering medical freshmen rose sharply, leaving candidates with solid, yet unremarkable numbers—still indicative of success in medical school—increasingly likely to face rejection; many of these students reapplied the next year, often to a wider sample of medical schools[19]. With more applicants filing more applications than ever before, the problem became deciding which students were *most* qualified, not which students were qualified enough. Though increased diversity in applicant ranks was a healthy step for the future of medicine, the new heterogeneity of applicants compounded the problems of the selection process even further; desirable minority applicants did not always present the same set of qualifications as the desirable white male applicants the system had developed around[20]. Furthermore, the precise definition of the "most qualified" student of any sort was not universally agreed upon. Many commentators argued students with lower objective scores but a strong commitment to practice in underserved rural or inner-city communities were more worthy than students with high grades, high MCAT scores, and a dearth of medical motivation[21]. Those concerned with the shrinking number of primary care physicians pointed to the tendency for the highest baccalaureate achievers to pursue research or

[19] "Undergraduate Medical Education", p. 1340-1.
[20] For a comprehensive account of demographic trends in medical students, see Davis G. Johnson, *Physicians in the Making: Personal, Academic, and Socioeconomic Characteristics of Medical Students from 1950 to 2000* (San Francisco, Jossey-Bass Inc., 1983), p. 12-160. For an examination of social and intellectual changes in the student population, see , e.g., Rodney M. Coe, et al, "The 'New' Medical Student: Another View" *JME* 52, no. 2 (February 1977), p. 89-98, or, for a slightly earlier history, Naomi Rogers, "'Caution: the AMA May Be Dangerous to Your Health': The Student Health Organizations (SHO) and American Medicine, 1965-1970" *Radical History Review* 80 (2001), p. 5-34.
[21] See, e.g., Paul G. Rogers, "Congressional Perspectives on Government and Quality of Medical Education" *JME* 51, no. 1 (Jan. 1976), p. 3-6, and Therman E. Evans, "Reverse Discrimination" in the same issue, p. 80-82.

specialty practice[22]. Finding solutions to localized manpower shortages demanded taking chances on students whose high motivation balanced out middling grades and MCAT scores. Even those agreeing on the importance of motivation had no way to quantize this information; the search for a different kind of ideal student called for psychological or personality testing to replace achievement and aptitude tests[23]. The surplus of intellectually capable students pushed medical educators to look further into students' capacity for medical practice, rather than merely the ability to graduate with an M.D. The existing metrics, especially the MCAT, proved incapable of providing a longer window into the future, making the test seem increasingly irrelevant to the needs of medical education.

As the admissions process began to buckle under the weight of excess applicants, calls for change grew louder, often proposing drastic solutions to a difficult problem. Prior to the *Bakke* decision, some commentators favored expanding quotas beyond race and home state, dividing one large pool of applicants into several smaller, more manageable groups by gender, age, or socioeconomic standing[24]. Though it promised to make competition in the more populous categories (particularly affluent white male biology and chemistry majors) rather cutthroat, quota expansion guaranteed a diverse student population. Other proposals retained the MCAT and other methods of weeding

---

[22] Harry Perlstadt, "MCAT: A Gate in Admissions and Internship Placement" *JME* 50, no. 1 (Jan. 1975), p. 78-81, and Harrison G. Gough, "Some Predictive Implications of Premedical Scientific Competence and Preferences" *JME* 53, no. 4 (Apr. 1978), p. 291-300. For more on the generalist physician response to the rise of specialization, see Robert Bartz, "Generalists First: The Movement to Refashion General Practice in Post-World War II America" (unpublished Ph.D. dissertation, UC San Francisco, 2006).

[23] See, e.g., John V. Haley and Melvin J. Lerner, "The Characteristics and Performance of Medical Students During Preclinical Training" *JME* 47, no. 7 (Jul. 1972), p. 446-52, and Penelope Kegel-Flom, "Predicting Supervisor, Peer, and Self Ratings of Intern Performance" *JME* 50, no. 8 (Aug. 1975), p. 812-5.

[24] All of the following ideas are reviewed in Robert A. Green and Davis G. Johnson, "Solutions Needed for Medical School Admissions Problems" *JME* 47, no. 12 (Dec. 1972) p. 974-6. All public medical schools and a number of private institutions continue to reserve a bulk of their seats for in-state residents, often as mandated by state law.

out weak students, but relied on a new process to decide who could become a physician, either by following the European and South American model of accepting all qualified applicants and cutting weaker medical students later down the line, or by installing a random lottery assigning seats to acceptable students[25]. The lottery suggests a complete lack of faith in the existing system's ability to identify the best potential doctors, rather than simply the best potential medical students. Using the MCAT to predict clinical success seemed no better or less arbitrary than a random drawing; furthermore, a student rejected for not having high enough scores would hear a much different, far more negative message than one rejected by random chance. No schools seem to have adopted any of these methods, but their mere consideration speaks to the severity of the admissions problem, and the absence of any simple solution to the problem.

Faced with the threat of more applicants filing even more applications for years to come, the AAMC called together a conference to overhaul the entire admissions process. The Medical College Admissions Assessment Program proposed several changes in admissions, but the most dramatic and lasting of these affected the standardized test[26]. Four regional conferences met in 1973, bringing together a variety of representatives, including deans, students, faculty, and undergraduate advisors[27]. Each group proposed its

---

[25] Ibid. Discussions of the lottery process can also be found in N.L. Eckhert and E.J. Cronin, "Diversity in Membership in Medical School Admissions Committees" *JME* 59, no. 8 (Aug. 1984) p. 635, and Emily Mumford, "Protecting the Public, the Profession, and the Applicant" *JME* 59, no. 12 (Dec. 1984) p. 971. Though most references to lotteries seem to be as much for rhetorical effect as anything, Eckhert and Cronin note that the Florida State Legislature considered a bill to install them in 1979.

[26] See "Final Report of the AAMC National Task Force" from note 17, and Nancy Cole, et al, *A Plan for the Development of a Medical College Admissions Assessment Program* (Iowa City, American College Testing Program, 1973) in the same box, Folder 3, which include expository accounts of the motives and intentions of the MCAAP conferences and programs. Other ideas proposed included standardized recommendation and evaluation forms, in order to enhance the information that could be drawn from them.

[27] *Position Papers for Regional Conferences* (Washington, DC, AAMC, 1973), in "Medical College Admissions Assessment Program: Box 1", Folder 1.

own suggestions, but nearly every group recommended replacing the MCAT[28]. The concern that potentially excellent doctors were losing their educational opportunities to students who were only impressive on paper led to a consensus that change was imperative. There was little agreement, however, on how precisely to fix the situation. Some groups wished to simply repair the MCAT by removing the most troublesome sections, while others hoped to replace the MCAT with an expanded battery of more effective measurements than the current test was capable of making. Proposals sometimes clashed: the northeast's medical education group and student representatives both called for "a test that measures only achievement," while students from the central region recommended introducing an aptitude-only test[29]. The only specific conclusion nearing full consensus was the desire to remove General Information, "The weakest area of the MCAT," still carrying a high risk of bias and low predictive validity[30]. Whatever the MCAT's replacement ended up becoming, it had to offer a new way to compare the glut of well-qualified applicants that was both fair and accurate. The original bar set by the MCAT could not simply be raised higher to accommodate the applicant boom; this fact led to a few of the more drastic ideas winning out over more conservative proposals.

The unlamented death of General Information speaks to the necessities driving the eventual overhaul of the MCAT. No one had ever intended for General Information to participate in the final decision between acceptance and rejection; it was actually something of a luxury, a way to encourage liberal study from a manageable number of

[28] The lone exception of the twenty position papers came from the Western region's student representatives, who offered possible suggestions for a revised test, but ultimately decided, "The question as to whether the MCAT ought to be eliminated or revised should be determined by admissions officers." Ibid, p. 229.
[29] Ibid, p. 83, p. 111, and p. 52, respectively. The latter organization further endorsed the use of personality measurement, while opposing multiple science discipline scores (p. 53); the southern region's student affairs group worried about the reliability of non-cognitive measures, but was one of several groups to endorse individual scores for each science discipline (p. 140-44).
[30] Ibid, p. 149.

applicants[31]. As that number became increasingly harder to manage, General Information became useless, disliked by most examinees and unused by most admissions staff. The replacement test needed not only to highlight potential failures, but also to provide a rough estimate of how the likely successes would stack up against each other. Promoting non-scientific knowledge, or any other purpose for the MCAT, was of minor importance.

The definition of "success" as estimated desperately needed revision. Almost every matriculant would earn an M.D., but some students would prove better than others, and some were likely to become better doctors than others; these groups did not necessarily commingle. A 1974 Duke Medical School survey found nearly three-quarters of students earning honors marks in clinical education had not received a single honors grade throughout the basic science curriculum[32]. More disturbingly, the inverse was also true: most students with honors in basic science coursework had unremarkable clerkships. Beyond this, the performance of many science honors students around patients was "subtly disturbing because they appeared to be uninterested in patients as people."[33] The first half of medical school seemed easier to predict, because it closely resembled prior education; undergraduate science GPA proved the single best predictor of first- and second-year GPA, and the MCAT the best predictor of the first stage of the National Boards, dominated by basic science. However, the noted disconnection between the two halves of medical school furthered the faults of an admissions process designed to predict the half unrelated to clinical performance. "Success" in medical school needed to move

[31] See Erdmann, "Separating Wheat from Chaff." There is also a connection here to the concept of the ever-elusive "well-rounded" student, an idea at all stages of admissions used to discourage overly bookish students from dominating the population. On the undergraduate level, see especially Karabel, p. 139-247.
[32] John M. Rhoads, "Motivation, Medical School Admissions, and Student Performance" *JME* 49, no. 12 (Dec. 1974), p. 1119-1127.
[33] Ibid, p. 1121.

beyond the absence of failure, and beyond scholastic performance itself, to reflect the actual purpose of medical education.

The MCAAP conferences' final report recommended a new test to replace the existing MCAT as soon as possible. The report endorsed further research into replacing other aspects of admissions, in order to create a more cohesive, efficient process, but focused most of its specific recommendations on developing "achievement measures designed purposely to determine the applicant's preparation for medical education."[34] Both aptitude and achievement measures were included, but both required improved relevance. The MCAT—under the AAMC's control, and having already helped accomplish its original goal—seemed the ideal vehicle to deliver new and better pieces of information on the future potential of applicants. The existing MCAT survived until the new test was ready in 1977; even though a consensus for change had been reached, no evidence suggests medical schools approached the lame-duck MCAT any differently than before[35]. Many schools had reservations about the MCAT, but had little else available to help make their difficult decisions. Developing a new, better metric was an urgent demand.

LOOKING LIKE MEDICAL SCHOOL

To offer a finer definition of success, the MCAT's replacement seemed obligated to more closely resemble the medical school curriculum. To develop this test, the AAMC

---

[34] "Final Report", p. 18. Other measures endorsed in the Final Report (to varying degrees of actual execution) included standardized letters of recommendation, a structured outline for personal statements, and an examination into use of the interview.

[35] Average matriculant MCAT scores continued to rise between the final report's recommendation in 1973 and the introduction of the New MCAT in 1977. See Travis L. Gordon and Davis G. Johnson, "Study of U.S. Medical School Applicants, 1976-77" *JME* 53, no. 11 (Nov. 1978), p. 873-897.

turned to the American Institute for Research (AIR) of Palo Alto, California, architects of the National Board of Medical Examiners test revisions in 1959[36]. AIR prepared a survey of science topics for possible inclusion, and sent it in two directions: medical school faculty assessed each topic's relevance to their curricula, while undergraduate science faculty reported whether their schools' premedical core courses adequately covered the material[37]. Topics found both medically relevant and available to most undergraduates comprised a content outline sent to professors to develop questions, with the most relevant topics allotted the most questions per test[38]. This outline represent an attempt to codify the most important parts of premedical education, without requiring students to go out of their way to prepare for the MCAT. Each of the changes that comprised the New MCAT attempted to make the test more relevant, more efficient, and more predictive of later performance; though the latter remained unknown for several years, the gains in the former two categories offered room for optimism.

The new test comprised four sections with six scores, more than doubling the test day to accommodate the added metrics. Science questions split into Knowledge and Problems sections; the latter section, comprised of short passages followed by sets of three related questions, received a separate "Science Problems" score intended to measure an examinee's critical thinking skills[39]. This new feature tried to analyze a new set of knowledge, one closer to what modern medicine needed. Separate scores were

---

[36] "Introduction to MCAAP 1976 Admissions Tests", p. 3, in "Medical College Admissions Assessment Program: Box 1", Folder 14. The NBME changes were also directed at more clinical relevance.
[37] The original survey is not in the AAMC Archives, though a 1981 follow-up modeled on the original is included as "MCAT Survey of Premedical Science Course Content" in "MCAT: Box 1", Box 501982734, Folders 19 and 20.
[38] Again, the original is not included, though reference to it is made in the final report of the 1981 surveys in "Review of Science Content Specifications for the Medical College Admission Test" in "MCAT: Box 2", Box 501982734, Folders 3-7.
[39] See "Introduction to MCAAP 1976 Admissions Tests" from note 36, and the public announcement in John A.D. Cooper, "The New Medical College Admissions Test" *JME* 52, no. 1 (Jan. 1977), p. 77.

derived from both Knowledge and Problems section in Biology, Chemistry, and Physics, thus turning the old test's single science score into a four-score diagnostic[40]. The verbal section became Skills Analysis: Reading, exchanging analogies and synonyms for comprehension questions on medically related passages; similarly related charts, graphs, and tables formed the basis of the new Skills Analysis: Quantitative subtest[41]. General verbal and mathematical aptitude seemed imprecise predictors of how an examinee would handle specialized medical data; the Skills Analysis sections thus attempted to make even basic aptitude skills closer in tune with medical demands. A more direct connection to medicine centered all six scores, and lay behind the heavy new emphasis on science.

Efforts were made to distance this new test from the old MCAT. As late as 1975, documents still suggested that the new test would drop the MCAT name altogether, in favor of MCAAP[42]. Thought it would eventually retain the Medical College Admissions Test, all AAMC mentions of the test for several years referred to it as the "New MCAT", the N always capitalized[43]. This rebranding emphasized the desire to separate the new test from the old, furthering the distance already established by the New MCAT's structural overhaul. The SAT-style scoring system yielded to a curved, fifteen-point scale intended

---

[40] Ibid. See also Cole, et al from note 26, one of many places discussing the fear that an omnibus score gave the same marks to students with average scores across the board and those whose strengths in biology were balanced out by a weakness in physics (p. 4-6). This fear was always discussed in terms of those strong in the most apparently relevant science (biology) and weak in the least important (physics).

[41] See note 39.

[42] "Introduction to MCAAP 1976 Admissions Tests", which was published the year before its title would suggest.

[43] This is most prominent in early discussions, such as Cooper (note 39), and Mary Littlemeyer, ed., *New MCAT Student Manual* (Washington D.C., AAMC, 1977), but the capital N can be seen into the mid-1980s, as in Anne E. Crowley, et al, "Undergraduate Medical Education" *JAMA* 256, no. 12 (Sept, 26, 1986) p. 1563.

to minimize irrelevant score differences[44]. In addition to discouraging comparisons between old and "New" MCAT scores, the new scale intended to make score gaps meaningful at all points, whether a seven versus nine or a twelve versus fourteen[45]. The incompatibility of the old and "New" MCAT's allowed for the possibility of using the test in different ways. The AAMC organized a series of evaluative studies "directed at identifying as precisely as possible the constraints that should be placed on the scores reported," though it acknowledged the necessary delay in finding these results[46]. Data on first-year medical school grades of 1977 examinees became available in mid-1979, after five administrations of the test had been given. It would take even longer to analyze the New MCAT's correlation with the second half of medical school.

Given the lapse between the first administrations of the new test and the availability of the first correlative data, a certain amount of blind faith in the temporarily unknown usefulness of the MCAT was required. The validity of a new test could only be demonstrated through experience—perhaps why one MCAAP conference participant warned, "the rather long and somewhat bloody past history of other 'struggles' to revise the MCAT...is littered with the bones of effort to develop new and better tests."[47] It still remained possible that the new test would prove a weaker metric than the old one, though nobody in the AAMC seems to have publicly acknowledged this, or developed a contingency plan. By streamlining the test's content into a more relevant framework,

---

[44] *New Medical College Admission Test Interpretive Manual* (Washington D.C., AAMC, 1977), found in "MCAT: Box 1", Folder 12.

[45] The *Interpretive Manual* explained explained the test had an error range of plus or minus one point, so the difference between a nine and an eight was not quite so authoritative. Furthermore, the rigid content outlines and score scales attempted to make an eight scored in 1977 indicate the same level of knowledge as an eight scored in 1987. Evidence as to whether this was achieved is not readily available in the archives, and would at any rate require a dizzying amount of statistical analysis.

[46] Cooper, p. 77.

[47] Quote from Woodrow W. Morris of the Central Regional Group on Student Affairs, *Position Papers* p. 20.

better predictive validity promised to naturally follow. At any rate, however, the need to respond to the new state of medicine made changing the MCAT so necessary that the AAMC could not afford to move too cautiously.

ADAPTING TO CHANGE

Though no iteration of the test survived longer than eighteen years, and all but two revisions included major restructuring, the 1977 New MCAT represents a uniquely dramatic departure[48]. Beyond changes in form and content, the purpose of the MCAT moved past the failure-predictive intent of the original; the New MCAT was designed to do more than the old test ever could. Although the New MCAT's ability to forecast success remained unknown, the tighter connection to medical curricula and improved diagnostic function were noticeable from the first administration in April 1977. Both real and promised changes attempted to respond to fast-moving expansions of scientific and medical understanding. A new flexibility bound the MCAT to follow developing trends more closely.

It was first necessary to catch up the MCAT with the three decades of progress since its last major revision. Most of the material covered by the traditional premedical core remained fairly static between 1946 and 1976, but this period witnessed a few key developments that trickled down to basic undergraduate instruction, including the central dogma of molecular biology following the discovery of DNA[49]. AIR's content surveys

---

[48] McGaghie, p. 3. The 1962 revision primarily replaced Understanding Modern Society with General Information, while the revision beginning in 2007 will retain the same structure as the current exam, only with fewer questions per subtest, as the MCAT moves from paper-and-pencil to computer-based examination.

[49] There is a question on DNA included in the test included in Gladstone, et al (p. 24). It is certain that the 1976 science section did not test science as it stood in 1946, of course, but the old test never seems to have set a limit on what would or would not be asked, or when it would be asked. For a discussion of the central

speak to a newfound desire to keep the MCAT better tuned to science; reviewing the entire science content of the test every few years enabled the test to remain within the shifting limits of what information an aspiring doctor had to know. Since the first content surveys, a few items were even removed from the "Not to be Tested" category and introduced as testable content. The expansion of population genetics brought the Hardy-Weinberg equation onto the list of viable Biology topics, while the introduction and popularization of MRI technology seems to have made proton NMR a valid topic[50]. The pretesting and bias-checking required of MCAT questions forced the test to be less readily altered than curricula at either the undergraduate or medical school levels, but repeated surveys allowed for a response to changing knowledge requirements. This response seems a necessary step to prevent the test from losing all of the newfound relevance it attempted to claim.

Beyond individual topics, the New MCAT's focused pursuit on medical relevance itself reflects a change in what medical institutions needed from the applicant pool. As increasing legal and financial interventions in patient care focused medical aims onto ends and results, constant technological innovations expanded the complexity of medical knowledge, while reducing dependence on the basic skills of old. As medicine became, in the words of David Rothman, "Powerful and impersonal, a more or less efficient interaction between strangers," scientific competence became far more important than a

---

dogma from the man who won a Nobel Prize for helping to formulate it, see Francis Crick, "Central Dogma of Molecular Biology" *Nature* 227 (1970) p. 561-3. The dogma has since been expanded and amended to accommodate a number of unforeseen discoveries, the importance of which are far outside the scope of this paper.

[50] Comparisons made between items listed as "Not to be Tested" in the "Review of Science Content Specifications" from September 1981 mentioned in note 38 (Folder 3), and in Appendix C of the most recent *MCAT Interpretive Manual* (Washington, D.C., AAMC, 2005). Nuclear magnetic resonance was not untouched by the 1977 outline, but it appears to be much more important on the current test.

well-rounded education[51]. Medical school admissions had already begun choosing candidates with these new needs in mind, even before the applicant boom or the MCAAP. Some of the New MCAT's changes attempted only to work around how admissions committees had been using the MCAT for years, by adding information that had been improperly inferred by medical schools in the past.

THE SAME AS BEFORE

These improper inferences were not borne out of malice, or even ignorance. Finding a practical method of reviewing thousands of applicants for hundreds (sometimes dozens) of seats often meant cutting through the intended purpose of the MCAT to find the most efficient way of using it. The ultimate meaning of the MCAT depends on how admissions committees used scores to judge applicants. Even though the redesigned test attempted to provide new information and tackle new problems, the demands of admissions dampened some of the New MCAT's impact. Some schools seem to have used the New MCAT as they used its predecessor, making no adjustments beyond rescaling formulas to fit a different set of scores.

Though not every school had firm guidelines in considering an applicant's quantitative record, some used developing computer technology to produce hard formulas for sorting through applicants. Few schools were willing to publicly display their selection methods, but a handful of pioneering schools discussed how they created and revised their admission formulas. UC Irvine searched for the top few hundred students as weighed by a percentile ranking of GPA and MCAT Science and Quantitative scores,

---

[51] David J. Rothman, *Strangers at the Bedside: A History of How Law and Bioethics Transformed Medical Decision Making* (New York, Basic Books, 1991) p. 262.

"conserving its clerical and professional resources for carefully considering a reasonable number of applicants."[52] In reducing the time consumed in processing applications, Irvine cut out two of the four scores of the old MCAT; though Verbal and General Information may have played a role judging students from within the reasonable number, neither meant as much as Science. This lack of meaning derived solely from the weak connection between UCI medical students' Verbal and General Information scores and later performance; since science GPA best predicted medical GPA, and MCAT Science scores best predicted NBME Part I scores, these factors carried more weight, at Irvine and elsewhere. Irvine still considered promising students outside the computer-selected range, and suggested that cutting procedural tedium allowed staff to spent more time dealing with "often neglected-issues of admissions policy and educational goals and philosophy."[53] After observing the predictive capabilities of the New MCAT, Irvine was among the first schools to develop an updated admissions formula; Science Problems and Skills Analysis: Quantitative proved valid enough to enter the formula, while the other four scores were discarded[54]. Considering the features of the New MCAT would have required the time that the old formula had saved, so Irvine chose instead to rework the existing model with the closest analogues of the old variables. It is likely this issue extended beyond UC Irvine and other schools with formulas; dealing with so many

---

[52] Frederick L. McGuire, "Fifteen Years of Predicting Medical Student Performance" *JME* 52, no. 5 (May 1977) p. 416-417. Science scores received double the weight of Quantitative, and Science GPA. See also J. Hutchison Williams, et al, "A Computer-Assisted Admissions Process" in the same issue, p. 384-9, for a formula at The Ohio State University weighing raw scores rather than percentiles. Early versions of this formula lightly weighed Verbal and General Information scores, but ultimately Williams found more predictive success with GPA and MCAT Science alone. Nevertheless, this article is possibly the only one to explicitly mention consideration of General Information scores, if only for a year.

[53] McGuire, p. 417.

[54] McGuire, "The New MCAT and its Relationship to Student Performance—Year Two" *JME* 57, no. 1 (Jan. 1982) p. 60-61. This new formula also dropped the use of nonscience GPA altogether, leaving only science GPA and the two MCAT scores as variables.

applications in such little time demanded the occasional corner be cut, even by using admittedly imprecise metrics.

Nearly every study of the New MCAT conducted by a single institution focused on one of two issues of great importance in admissions: the test's predictive validity, or its use in separating applicants from each other. Early correlative analyses found the best predictions of future performance came from either Chemistry, Biology, or Science Problems, though the NBME Behavioral Science section uniquely corresponded to Skills Analysis: Reading[55]. Chemistry correlated most consistently with medical school performance, earning it the highest average weight from the slight majority of schools that admitted to differentially weighting scores[56]. This speaks only coincidentally to the importance of chemistry in medical school, given that Physics received the least weight. Despite being subject to the same content demands as the other science tests, Physics held the least predictive validity, and thus assumed dubious practical significance. This weighting also curbed the use of the MCAT as diagnostic; if two otherwise equal students presented alternating Chemistry and Physics scores, one high, one low, medical schools were likely to simply take the high Chemistry scorer over the high Physics scorer. Remedying weak knowledge in Physics was unnecessary, and remedying weak knowledge of Chemistry was unnecessarily risky, given that many other students

[55] See, e.g., Thomas J. Cullen, et al, "Predicting First-Quarter Test Scores from the New Medical College Admission Test" *JME* 55, no. 5 (May 1980), p. 393-8; Charles P. Freidman and William E. Bakewell, Jr., "Incremental Validity of the New MCAT" *JME* 55, no. 11 (Nov. 1980) p. 967; and C. Michael Brooks, et al, "Validity of the New MCAT for Predicting GPA and NBME Part I Examination Performance" *JME* 56, no. 9 (Sep. 1981) p. 767-9. The first study reaching past the first two years is Jan D. Carline, et al, "Predicting Performance During Clinical Years From the New Medical College Admission Test" *JME* 58, no. 1 (Jan. 1983) p. 18-25.

[56] Karen J. Mitchell, "Examination of MCAT Format and Content Specifications" in AAMC Archives "A.G. Swanson MCAT Papers", Box 501982516, Folder 6, p. A-2. Some schools (including Irvine) used Science Problems in the place of Chemistry, but many schools opted against using the Problems score at all, since its scoring overlapped with the three science discipline subtests.

presented scores more directly indicative of success. Formulas and weighting perpetuated the use of the MCAT as predictor of success, even before long-term research on the new test's predictive value became available.

Not every formula proved so exclusive, but each one sought a similarly limited meaning of the MCAT. A joint effort between Florida A&M and the Florida State University Medical School gave extra weight to GPA and MCAT Science and Quantitative scores, but also devised point scales for socioeconomic and demographic ratings, nonacademic qualifications, and even Myers-Briggs type indications[57]. Poorer rural students with strong sensing qualities thus had the ability to make up for underwhelming marks by presenting qualities associated with primary care physicians in underserved areas. Even this admirable effort, however, attempted to use admissions as an oracle, sifting given and created numbers into a vision of success or mediocrity. This function handled both old and New MCAT scores in the same way. Early results suggested the New MCAT made only marginal gains over the old test, though it proved effective enough to survive. Still an imprecise predictor, the New MCAT, particularly when used no differently than its predecessor, held over many of the same flaws and quirks.

IT DEPENDS ON WHO YOU'RE LOOKING AT

[57] Paul R. Elliott, "The Selection of Primary Care Physicians," presented at the Workshop for Undergraduate Education in Family Medicine, Kansas City, Missouri, May 12-14, 1975, ERIC Fiche # ED 134060. The program was designed to provide a guaranteed space at FSU for promising A&M underclassmen, who were then allowed to use their last two undergraduate years as they wished.

For many schools, the weight (and meaning) of the MCAT depended on the examinee. The need to diversify medical practice led medical schools to accept underrepresented minorities with below-average scores. This double standard became the means of dealing with a chronic score imbalance between white and minority students, though not without controversy and legal intervention. As low-scoring minorities proved generally successful in medical school, especially closing the gap in the clinical phase, the MCAT's predictive value suffered. With entire categories of people insufficiently described by the MCAT, questions emerged about whether the test ought to be used at all.

Before coping with the applicant boom, the biggest crisis in medical education was a disconnection between physician demographics and those of the population at large. Twelve percent of Americans were in a racial minority in 1970, compared to two to three percent of medical students, most of whom attended a small handful of schools led by the traditionally black medical schools at Howard University and Meharry Medical College[58]. A homogeneous student population threatened to exacerbate existing medical manpower disparities; the traditional pool of high-achieving, well-off white male medical students tended towards specialty practice in wealthy communities, flooding the suburbs with doctors at the expense of the inner cities. The AAMC responded in 1970 by initiating "Project '75", an attempt to promote medical diversity by pushing for a

[58] Discussed in "Undergraduate Medical Education" in the 1975 *JAMA* Education Issue (see note 13), p. 1339. The semantic difficulties of discussing race make it difficult to discuss a constant group of people on one side or another. For instance, the twelve percent minority population figure is taken solely from the 1970 Census's total of African-Americans, with other groups small enough to be statistically negligible. An influx of Hispanics and Asians during the period covered in this paper, however, altered the American composition. Further complicating the issue, Asian students perform roughly similar to white students, in medical school as elsewhere, and thus many later studies opted to lump whites and Asians into an "other" category weighed against the "underrepresented minorities" of black, Hispanic, and Native Americans, who do not achieve at the same scholastic level on average. In discussing racially-separated studies, I have attempted to remain as close as possible to the classificatory language used by each paper's authors.

minority student population of twelve percent by 1975. This figure was never met, with minority totals peaking at slightly less than ten percent of medical students in 1974[59]. Since these growing ranks seemed the best hope of medicine to extend into underserved communities, Project '75 took extra steps to promote medicine to promising undergraduates; medical schools went further to retain these students, by offering remedial education programs and offering a more flexible curriculum. In addition to permitting those who needed to finish in five years to do so, advanced students could make it through in three. These extra measures acknowledged an educational gap between white and minority students, and took drastic action to close it before it could endanger lives at the clinical stage.

The desire to diversify medicine led medical schools to treat minority status as a balancing factor against lower grades and test scores. Alongside a persistent GPA gap, black examinees score below the five-hundred point median on all the old MCAT subtests, even as white examinees pushed the overall mean above six hundred for some subtests[60]. Even with the elimination of General Information, this trend continued on the New MCAT; white examinees in 1978 averaged above an eight on all six subtests, while means for black examinees ranged from 4.86 in Quantitative to 5.55 in Physics[61]. Even at

---

[59] For an overview of Project '75, and a mission statement that eventually built into the 1990s effort "Project 3000 by 2000", see Robert G. Petersdorf, et al, "Minorities in Medicine: Past, Present, and Future" *AM* 65, no. 11 (Nov. 1989) p. 663-70. African-Americans made up the bulk of these totals, representing 7.5% of all entering medical students in 1974.

[60] See, e.g., 1970s *JAMA* Education issue "Undergraduate Medical Education" sections, and W.F. Dube and Davis G. Johnson, "Medical School Applicants, 1972-73" *JME* 49, no. 9 (Sep. 1974), p. 849-69, for means, medians, and breakdowns by category.

[61] *New Medical College Admission Test Percentile Rank Ranges for New MCAT Areas of Assessment: 1978 Summary of Score Distribution* (Washington, D.C., AAMC Division of Educational Measurement and Research, 1978) p. 10-11. Published percentile rank ranges are available between 1978 and 1985, and various editions includes breakdowns by gender (showing a slight advantage for females on the Reading subtest, and a slight advantage for males on the other five sections) and by state. California, Idaho, Minnesota, and Utah proved consistently high-scoring, while the District of Columbia and U.S. Territories (including Puerto Rico) ranked at the bottom.

the peak of Project '75, the bulk of medical school applicants were white (and male); thus, the overall average score stayed close to the Caucasian mean. Color-blind admissions policies would have placed an average underrepresented minority student in the bottom half of the applicant pool, below several thousand white and Asian students. Increasing diversity meant taking minority students with lower grades and scores than were acceptable within the rest of the pool.

Admissions officers used a variety of means to achieve this end. MCAT subtest weights proved malleable following research showing aptitude-based verbal/reading scores bore a higher correlation with medical school performance for minority students than science achievement scores[62]. Some degree of demonstrated intelligence—in this case, verbal acumen—remained useful in spotlighting students in spite of a larger educational gap. Other schools compensated for score weaknesses of even less consequence; commonwealth Puerto Ricans tested poorly in Skills Analysis: Reading, but the island's three medical schools taught much of their curriculum in Spanish, and thus largely disregarded Reading scores[63]. Most schools, however, seem to have

---

[62] A series of studies exploring minority medical school admissions were commissioned by the Department of Health, Education, and Welfare and conducted by the RAND Corporation, including John E. Rolph, et al, *Predicting Minority and Majority Performance on the National Board Exams* (Santa Monica, Cal., RAND, 1978), available as ERIC Fiche # ED 181075, and Albert P. Williams, et al, *Factors Affecting Medical School Admissions Decisions for Minority and Majority Applicants: A Comparative Study of Ten Schools* (Santa Monica, RAND, 1979), available as ERIC # ED 186460. The latter found that all ten schools under study used some special procedure for handling minority applicants, and gave rise to a later talk by Williams, "Predicting Performance in the Medical School Continuum: Toward Better Use of Conventional Measures", delivered at Duke University, April 1980, available as ERIC # ED 202402. See also Charles Odegaard, *Minorities in Medicine from Receptive Passivity to Positive Action* (New York, Josiah Macy Jr. Foundation, 1977), which recommended evaluating minority students' self-concept, ability to cope with racism, and potential for community leadership instead of using cognitive measurements. Odegaard, from the University of Washington's medical school, was frequently mentioned in admissions discussions, but it is not known how much his ideas came into practice outside of his own school.

[63] Commonwealth Puerto Rican examinees' average Reading scores in the eight available volumes of *Percentile Rank Ranges* varied from a high of 3.81 in 1979 to a low of 3.04 in 1981. The Ponce School of Medicine and the medical schools of the University of Puerto Rico and Universidad Central del Caribe required bilingual ability as a prerequisite in their yearly entries in the AAMC's *Medical School*

evaluated the same information, but set two different thresholds[64]. Underrepresented minorities needed to meet a lower GPA and/or MCAT standard to warrant further consideration at schools where the ultimate decision relied on interviews rather than scores. Some schools set up two separate competitions altogether, establishing a quota of seats reserved each year for minority students.

Allan Bakke, a white applicant denied a place at the UC Davis Medical School partially due to such a quota, challenged the constitutionality of all race-factored admissions. Davis saved sixteen spaces in each class of one hundred for minority students. Selecting the other eighty-four seats involved a benchmark formula factoring in MCAT scores, extracurriculars, and interview results, among others; Bakke's scores just missed consideration in the regular pool, and he was not considered to fill unoccupied spots reserved for minorities. The Supreme Court eventually split on the issue, leaving the holding of *Regents v. Bakke* in the hands of Justice Lewis Powell, who ruled hard racial quotas illegal, but did state the goal of enhancing student diversity was "sufficiently compelling to justify consideration of race in admissions decisions under some circumstances."[65] Establishing two different applicant tracks proved unconstitutional, but many systems factoring in race were still valid; a system like the one used by Florida A&M and Florida State could continue to give extra points to minority applicants likely to practice in underserved communities. Minority applications

---

Admissions Requirements guide (Washington, D.C., AAMC, published yearly beginning in 1949), hereafter referred to as *MSAR*.

[64] *Factors Affecting Medical School Admissions Decisions* (see note 63). See also William E. Sedlacek and Dario O. Prieto, "Predicting Minority Students' Success in Medical School" *AM* 65, no. 3 (Mar. 1990) p. 161-6.

[65] *Regents of the University of California v. Bakke*, 438 U.S. 265, p. 267. Though the AAMC did file a brief with the Supreme Court supporting preferential admissions, *Bakke*'s significance extends far beyond the scope of this paper. For a dedicated discussion of the case's impact, see Howard Ball, *The Bakke Case: Race, Education, and Affirmative Action* (Lawrence, Kans., University of Kansas, 2000).

fell after *Bakke*, as did the total number of applicants, but the case established that superior quantitative results were not the lone route to acceptance[66].

By the time of the *Bakke* decision, many minorities accepted through initial Project '75 efforts had become physicians, performing far better than their scores predicted, thus weakening the usefulness of the test. Minority students proved slightly more likely to face academic difficulty, but retention efforts enabled all ethnicities to boast graduation rates above ninety percent[67]. Educational gaps persisted in the first half of medical school, but shrank during clerkships and post-graduate residencies[68]. Overall, these performances far exceeded the high risk of failure predicted by taking in so many students with below-average MCAT scores. Thus, not only did the test's desired function of predicting success prove inadequate, its original ability to predict failure weakened when analyzing non-white students. Test scores could be manipulated to better work with minority admissions, but doing so removed any inherent meaning from the readily malleable scores. An observer quoted during the MCAAP conferences summed up the dubious meaning of the MCAT by suggesting the minority situation was akin to testing

---

[66] C.H. William Ruhe, "Recent Events of Special Interest to Medical Education" *JAMA* 243, no. 9 (7 Mar. 1980) p. 842-6.

[67] "Undergraduate Medical Education" *JAMA* 243, no. 9 (7 Mar. 1980) p. 849-66.

[68] Studies did observe a tendency for students with academic difficulty to continue having problems, such as Leonard M. Rosenfeld, "Delays in Completing Medical School: Predictors and Outcomes" *Teaching and Learning in Medicine* 4, no. 3 (Autumn 1992) p. 162-7. But other studies point to the lack of connection between the two halves of medical school; in addition to Rhoads (note 32), see Gang Xu, "Longitudinal Comparison of the Academic Performances of Asian-American and White Medical Students" *AM* 68, no. 1 (Jan. 1993) p. 62-8, and Doug Campos, et al, "Performances of Underrepresented-Minority Students at the University of Arizona College of Medicine, 1987-1991" *AM* 68, no. 7 (Jul. 1993) p. 577-81. Another angle on this topic is offered by Thomas J. Stachnik and Ronald C. Simons, "A Comparison of M.D. and D.O. Student Performance" *JME* 52, no. 11 (Nov. 1977), p. 920-5, which noted no consistent difference between the two in the unique dual program of Michigan State, even though D.O. students tended to enter with weaker academic credentials. For a more anecdotal account, see the discussion of a troubled student turned competent doctor in Herbert Schapiro and Robert M. McCombs, "The Disadvantaged Student" *JME* 54, no. 8 (Aug. 1979) p. 672-3.

"a rubber band. The test may tell you how long it is now, but not how far it will stretch."[69]

THE POWER OF A NUMBER

Admissions officers knew few specifics of the New MCAT; physicians involved in student selection had taken its predecessor years before. Few took the opportunity to examine what made the New MCAT so new, or to understand the scoring process. Thus, the test's variable meaning was of minor concern, so long as predictive validities remained in order. The MCAT's greatest practical use had less to do with any imbued meaning, and more with its firm quantitative value. Most admissions variables, even GPA, were frustratingly subjective, leaving the chance that subpar students could sneak in over students with better potential. The MCAT provided a definitive gradation of all examinees; though the meaning of scores was imprecise and often fluctuated from school to school, those scores could nonetheless be assigned an objective meaning.

Though much of the MCAT's relevance depended on how well it correlated with medical school performance, the statistics seldom showed more than a moderate connection. Early analyses demonstrated some improvement in predictive validity between the old test and the new, but even with these gains, the MCAT rarely accounted for more than about twenty-five percent of the variability of NBME Part I scores, and a smaller proportion of first- and second-year grade variability[70]. Combining premedical

---

[69] *Position Papers for Regional Conferences*, p. 127. The source of this quote is not mentioned.

[70] See the studies cited in note 55, and also McGuire, "The New MCAT and Medical Student Performance" *JME* 55, no. 5 (May 1980), p. 405-8, and Randolph E. Sarnacki, "The Predictive Value of the Premedical Grade-Point Average" *JME* 57, no. 3 (Mar. 1982), p. 163-9. Squaring the statistical correlation between two factors produces the percent of variability explained by the connection between them; a 50% correlation thus explains 25% of variability, a 70% correlation explains 49% of variability, and so on.

GPA with MCAT scores proved a better predictor of both medical grades and NBME scores, but even so, the majority of difference in student performance could not be explained by the traditional criteria. High scorers tended to do better in medical school, but not to a degree that supported failsafe separation of students by quantitative measures. Supporters of the MCAT argued that unimpressive correlations were due to selection bias; since accepted students generally outscored those rejected, studies of MCAT validity looked only at the highest end of the scale, rather than the full spectrum. Students admitted to a BS/MD track at Northwestern University as high school seniors were required to take the MCAT midway through the program, but their scores did not affect their guaranteed place in medical school. These students demonstrated far more score variability than students admitted to Northwestern on the regular MD track. The combined GPA/MCAT correlation of these BS/MD students rose dramatically, accounting for as much as half of the variability in medical school measures[71]. Though data suggested the MCAT was effective across a wider body of subjects, most medical schools were not willing to accept low-scoring students to accommodate the test. None of the schools without MCAT requirements ever examined how their students scored on the test. Minority students offered some idea of performance correlations to lower MCAT scores, but there were not enough minorities to dramatically sway the results of the other ninety percent, who tended to be homogeneously high achieving. In theory, these students should have continued to perform equally, but a frustrating lack of uniformity

[71] Melton E. Golmon and Charles A. Berry, "Comparative Predictive Validity of the New MCAT Using Different Admissions Critieria" *JME* 56, no. 12 (Dec. 1981) p. 981-6. See also Kevin Hynes and Nathaniel Givner, "Restriction of Range Effects on the New MCAT's Predictive Validity" *JME* 56, no. 4 (Apr. 1981) p. 352-3.

emerged throughout medical school. Major differences between these students existed, but none of the traditional criteria were able to predict them.

A host of alternative methods promised to spot these unseen differences, or at least to improve on the MCAT's existing predictive validity; as long as the numbers seemed promising, few worried about what those numbers actually said. Noncognitive examinations were pushed during the early seventies, owing to the popular use of student interviews to gauge necessary factors like motivation and judgment that did not easily fit into a standardized test. Though the MCAAP intended to develop a centralized noncognitive examination, the project was "never implemented after an examination of its feasibility."[72] Nobody could agree on the specific noncognitive traits such a test needed to examine—traits found in the elusive concept of the ideal physician—and nobody knew how to effectively examine them. Unlike measurements of basic knowledge and aptitude, examining personality characteristics proved too unreliable, the death knell of any standardized assessment. Nevertheless, commentators continued to suggest that the unaccounted variance between students was the product of different mindsets[73]. Several schools found success analyzing selectivity at a student's undergraduate institution. The existing criteria supported the use of undergraduate selectivity; students with a GPA between 3.8 and 4.0 from low selectivity schools had the same average MCAT score as students with GPAs of 2.8 to 3.0 from the most selective schools[74]. Students from the most selective schools tended to do better than their GPAs

---

[72] "Review of the AAMC MCAT Program", p. 60, in AAMC Archives "A.G. Swanson MCAT Papers". Folder 5.

[73] See Sedlacek and Prieto, "Predicting Minority Students' Success" (see note 64), p. 165.

[74] Research factoring in undergraduate selectivity includes Robert F. Jones and Lori N. Adams, "The Relationship Between MCAT Science Scores and Undergraduate GPA" *JME* 58, no. 11 (Nov. 1983), p. 908-11, Doris A. Evans, et al, "Traditional Criteria as Predictors of Minority Student Success in Medical School" *JME* 50, no. 10 (Oct. 1975) p. 934-9, Terry T. Clapp and John C. Reid, "Institutional Selectivity as

predicted, supporting the theory that those successful at the most competitive institutions were likely to succeed at the next level. These selectivity rankings relied either on the Astin index or the classification in *Barron's Profile of American Colleges and Universities*, both of which based their rankings on the average SAT scores of matriculating students[75]. When used alongside or in place of the MCAT, these indices supplemented or replaced one standardized test with another. If unchecked by other measures, undergraduate selectivity ratings made it all but impossible for students from little-regarded institutions to enter medical schools with less than flawless grades. Since the indices worked well, however, they persisted in spite of concerns. Using the SAT did not particularly trouble admissions researchers; the SAT was essential in admissions for joint BS/MD programs, and a moderate correlation between the SAT and MCAT led to proposals to use the former test as a spotlight for potential premeds[76]. Though the SAT lacked all of the carefully structured science content of the MCAT, and came with the risk of pushing increasingly intense premedical competition to even earlier stages, it still held the potential to do the MCAT's job. Any method promising to beat the MCAT at its own game drew a share of support, especially those promising to shed light onto the second half of medical school.

Though many educators looked for a means to assess potential clinical acumen, such measures proved elusive and difficult to agree upon. One study evaluated the

---

a Predictor of Applicant Selection and Success in Medical School" *JME* 51, no. 10 (Oct. 1976) p. 850-2, and Rolph, et al, *Predicting Minority and Majority Performance on the National Board Exams* (see note 62).

[75] Alexander W. Astin, *Predicting Academic Performance in College: Selectivity Data for 2300 Colleges* (New York, Free Press, 1971) and *Profiles of American Colleges* (Hauppage, NY, Barron's Educational Series, twenty-six editions published since 1964).

[76] Vera B. Thurmond and Lloyd Lewis, "Correlations Between SAT Scores and MCAT Scores of Black Students in a Summer Program" *JME* 61, no. 8 (Aug. 1986) p. 640-3, and Jeremy R. Montague and Sister John Karen Frei, "A Twelve-Year Profile of Students' SAT Scores, GPAs, and MCAT Scores from a Small University's Premedical Program" *AM* 68, no. 4 (Apr. 1993) p. 306-8.

relationship of sixty-six different variables to clinical ability, but no obvious connections could be made; only eleven variables correlated with statistical significance, just four more than would be expected by chance[77]. One of the eleven was negative for the MCAT: high scores on the old Science subtest predicted sub-par clinical skills in the study; the strongest positive connections were found with emotional stability and Myers-Briggs Judgment typing, but even these were weakly correlated. Intuitively, low-scoring yet highly passionate students seemed the ideal candidates to fill much needed generalist positions in underserved communities, yet they came with the risk of falling behind in the curriculum before ever getting to display their clinical ability. The highest achievers presented little risk of failing two years of high-level scientific instruction, but there was no guarantee their classroom abilities would transfer to the emotional and interpersonal demands of patient care[78]. Ideal doctors could come from either group (as could incompetent doctors), but beyond the obvious outliers, measuring the quality of a doctor proved elusive. Clerkships were graded, and the latter two parts of the NBME focused on practical, clinical information, but assigning them an exact meaning proved no less difficult than doing so for the MCAT. Though many high achievers were cold around patients, they were also likelier to fill research positions that minimized their clinical

---

[77] Edward V. Turner, et al, "Predictors of Clinical Performance" *JME* 49, no. 4 (Apr. 1974) p. 338-42. Within the parameters used to calculate the correlations came the risk that about seven of the variables would exhibit statistically significant correlations due to chance error, rather than an actual connection. It is possible that none of the eleven variables would exhibit the same correlation over several followup studies (which were not performed), but it is likely that about four of those eleven factors do exhibit a relationship with clinical ability. Special thanks to Christiana Drake for explaining this somewhat counter-intuitive concept.

[78] This period did witness a number of attempts to bridge the gap between the basic science and clinical portions of medical school, beginning with the updated curriculum at Western Reserve University Medical School. See Greer Williams, *Western Reserve's Experiment in Medical Education and Its Outcome* (New York: Oxford University Press, 1980).

shortcomings[79]. The shrinking generalist population was a growing concern, but medical schools still needed to produce research physicians and specialty practitioners. These needs made the calculus of admissions still more challenging, and the lack of valid and detailed prediction methods more disconcerting.

Since admissions staff at each school manipulated MCAT numbers as they saw fit, a consensus approach to scores never developed. The AAMC abandoned plans to drop the MCAT point scale in favor of low, medium, and high ranges upon finding that scores of eight or nine could be unacceptable at one school and exemplary at another[80]. Even as applicant pools began to wane, elite medical schools received more exemplary scores than they could handle, but smaller schools still had to separate the acceptable from the unacceptable, utilizing the MCAT's original failure-predictive intent[81]. Medical schools might use the MCAT only in initial screening, to isolate the exemplary or eliminate the unacceptable from competition, or they could use scores throughout the process, even as a component of final selection after the interview[82]. The MCAT could be seen as a crudely effective way to lighten the burden on admissions officers, or as the single most important step in the process, depending on the school. Schools might even go past the AAMC's intentions; though summing the scores into a single composite was

---

[79] Mitchell J. Rosenholtz, "MCAT Scores and GPAs and Meeting Social Needs" *JME* 50, no. 2 (Feb. 1975), p. 214. Nonaccepted students examined in the same study expressed more desire to enter general practice, and less to enter research or specialty fields, than those who were accepted.

[80] Karen J. Mitchell, "Examination of MCAT Content and Format Specifications" in AAMC Archives "MCAT: Box 2" Box 501982734, Folder 21, and Mitchell, "Use of MCAT Data in Selecting Students for Admission to Medical School" *JME* 62, no. 11 (Nov. 1987) p. 871-9. Scores deemed "acceptable" ranged from four to ten, while "exemplary" scores ranged from eight to thirteen.

[81] This problem reemerged even more distinctly during the initial Ad Hoc Meetings on what became the next revision of the MCAT. William H. Luginbuhl, Dean at the University of Vermont Medical School, requested his name be removed from a 1986 report of final recommendations for a new test, recognizing he "was in a distinct minority at the meeting," attended by Deans from Johns Hopkins and Penn, and chaired by the Dean of UCLA's Medical School. Correspondence, Luginbuhl to August G. Swanson, 7 Mar. 1986, in AAMC Archives "A.G. Swanson MCAT Papers" Folder 5.

[82] See note 80.

not recommended, many schools did so to speed the process[83]. Since the New MCAT contained four science scores—one derived solely from questions factoring into the other three—composite scoring overextended the new scientific focus into a distinct disadvantage for examinees relying on Skills Analysis scores. The AAMC released the New MCAT without the data to support a specific, recommended use. Instead, schools were encouraged to perform institutional-level studies to determine how the MCAT best fit their needs[84]. These uses might cut corners, or expose flaws in the test's design, but demonstrably effective methods persisted. The absence of strict rules gave the MCAT flexibility, essential to functioning amidst the variety of different medical programs and philosophies throughout the country. The AAMC provided the numbers, and medical schools were relatively free to use those numbers as they wished.

The MCAT's greatest strength, however, was its status as the sole uniform point of comparison between almost every applicant. Other than high school seniors applying to BS/MD programs and students solely applying to the handful of schools without a strict MCAT requirement, every applicant took the MCAT, and all of them were weighed against each other[85]. Grades varied from institution to institution, as accusations of "grade

---

[83] Ibid.

[84] AAMC President John A.D. Cooper's editorial "The New Medical College Admissions Test" (see note 39) calls for smaller studies while outlining wider research designed to identify "as precisely as possible the constraints that should be placed on the use of the scores reported" (p. 77). However, Mitchell's *Use of MCAT Data in Admissions: A Guide for Medical School Admissions Officers and Faculty* (Washington, D.C., AAMC, 1987), published a decade later, is not particularly precise in its recommendations, instead offering a variety of scenarios used at various medical schools with some success. The book can be found in the AAMC Archives "MCAT: Box 2", Folder 12.

[85] The desire to ensure that students were compared on the same basis led to the requirement that all 1978-79 applicants needed to take the 1977 New MCAT. See *Announcement for the New MCAT* in AAMC Archives "MCAT: Box 1", Folder 11. The University of Rochester was the only school in the U.S. without a general requirement throughout the entire run of the New MCAT (other than the BS/MD-only University of Missouri-Kansas City); Cornell instituted an MCAT requirement in 1983, while Dartmouth, Columbia, Georgetown, and Case Western Reserve allowed for the occasional special exception from taking the test. Johns Hopkins dropped their requirement in 1986, a development that will be covered in the second half of

inflation" began flying, and even between different majors within the same institution; other than the basic premedical core, physics majors, biology majors, and English majors might have no other shared courses, offering a narrow overlap of common skills. Students with equal or incomparable qualifications otherwise might have a slight MCAT gap to act as a possible separator. Medical schools required personal statements and letters of recommendation, and frequently conducted interviews, but it was difficult to precisely rank this information. The difference between a twelve and a six offered seemingly definitive proof that the former student possessed greater knowledge in Chemistry or ability in Reading than the latter, and a stronger chance of medical success in some form. Most differences were not so obvious, but even with an admitted one-point margin of error, a series of elevens presented a better case than a series of tens, in a definitive way that impressive recommendations or statements could not provide. As a means of providing visible markers of separation within an often-indistinguishable applicant pool, the MCAT survived even while the meaning of those scores was up for debate.

DEFENDING A BLACK BOX

The MCAT, like all standardized tests, requires a difficult assumption be made about the information it provides: a student's knowledge can be reliably quantified through performance on an evaluation of necessarily limited time and scope. Since these examinations can help determine a student's admissions fate, the tests are carefully guarded, but this administrative secrecy complicates the heavy reliance on testing in

---

this history. Information taken from the 28th through 41st editions of the *MSAR*, eds. Verna E. Groo, Lynn Farley, and Vickie Wilson Ahari.

admissions. Both examinees and medical schools attach significant weight to the MCAT, but neither is privy to how the test's distinctions are made—even when those distinctions prove the difference between acceptance and rejection. A series of public attempts to open the black box of standardized testing placed little initial focus on the MCAT, but the unique demands of the AAMC fostered an aggressive, drawn-out campaign to shield the inner workings of the test from the demands of open public access.

The "truth-in-testing" movement began with the intent to affect the largest number of students possible. Thus, the undergraduate admissions tests, the SAT and ACT, drew most of the initial attention, since they affected the most lives, and possessed far greater visibility. Ralph Nader and Allan Nairn started their investigation at the Educational Testing Service in 1974, as New York's Public Interest Research Group lobbied the legislature for a "Truth-in-Testing" act[86]. Similar bills, which required test sponsors to provide the answers and solutions to all questions used to derive the score, soon came under discussion in several other state legislatures, and the U.S. House of Representatives[87]. In a post-Nixonian era, Americans were loath to blindly accept authority without accountability; truth-in-testing offered a check to any attempt by educational elites to exclude deserving outsiders[88]. The bill simultaneously expanded student rights while challenging monopolies with a history of racial and gender

---

[86] The ETS viewpoint on this story is covered in depth by Lemann, p. 218-232.

[87] One discussion of legislative movement came from one of the national bill's sponsors, who happened to be a Representative from New York: Ted Weiss, "National Truth-in-Testing Legislation" *Journal of Negro Education* 49, no. 3 (1980) p. 233-7.

[88] See James T. Patterson, *Restless Giant: The United States from Watergate to* Bush v. Gore (New York: Oxford University Press, 2005) p. 13-151 and Edward D. Berkowitz, *Something Happened: A Political and Cultural Overview of the Seventies* (New York: Columbia University Press, 2006) p. 1-83; 133-177 for discussions of the post-Vietnam, post-Watergate period giving way to the stagflation, deregulation, and eventually Reagan.

inequality. Medical schools and the MCAT were as guilty of these charges as any other level of education, and thus fell under the new law.

The AAMC proved the most aggressive opponent of the truth-in-testing movement, leaving the MCAT's status in New York in the air throughout the eighties. After New York passed the Truth-in-Testing bill, twenty of the twenty-six admissions testing programs threatened to stop administering their exams in the state, rather than comply with the law[89]. This tactic led some states (notably California) to dilute similar bills, while other state legislatures joined the House in scuttling their bills altogether, but New York would not compromise. Most other testing agencies backed down, complying with the new law or reaching private compromises, but the AAMC filed for and received a temporary injunction against the law in January 1980, days after the bill went into effect[90]. This injunction, centered on the MCAT's status as unpublished, copyrighted property of the AAMC, held for ten years, allowing the test to continue, but with a risk of uncertainty. Ten percent of all medical school applicants came from New York, and the law would have applied to out-of-state residents applying to any of New York's thirteen medical schools; though some of New York's schools took measures to protect students from an AAMC pullout, thousands of students took the MCAT without knowing whether their scores would matter[91]. ETS, ACT and other agencies adjusted to the bill, but the

---

[89] An admittedly biased recap of legislative activity is found in Erdmann, "Killing the Messenger" *JME* 55, no. 4 (Apr. 1980) p. 379-81.

[90] *Association of American Medical Colleges v. Hugh L. Carey, et al* 482 F. Supp. 1358, 22 Jan. 1980. The announcement mailed to students registered for the April 1980 examination (found in AAMC Archives "MCAT: Box 1", Folder 11), published after the ruling, still warned that the August administration might not be given in New York. Even taking the exam was not a guarantee that scores would be reported, since the AAMC could avoid having to release answers if no questions were used in scoring.

[91] New York had more applicants than more populous California during the seventies and eighties, and continues to have more medical schools (twelve allopathic, one osteopathic) than any other state. Examinee breakdowns by state are available in *Percentile Rank Ranges* volumes, while applicant breakdowns and individual school data are available in the annual *MSAR*. New York's size seems to have earned it the

AAMC fought it aggressively, revealing how crucial test secrecy was to the functioning of the New MCAT.

Truth-in-testing's mandated release of scored questions compromised the method used to ensure scores remained as comparable and consistent as possible across different forms and administrations. After establishing an initial question bank, each New MCAT administration used a three-form procedure; slightly more than half of all Saturday examinees received one form, and another quarter received the same form given to all Sunday test-takers. The third form, given to the remaining Saturday examinees, was identical to a form used two or three years before; the other two forms used one hundred eight previously-scored "anchor" items as well[92]. These used questions, imperiled by the New York law, served in score equating; since so much of each test tied back to previous administrations, scoring stayed consistent across administrations, enabling the AAMC to promise long-term score stability. Score equating was vital to support long-term research; if correlations between the MCAT and future performance seemed wildly inconsistent, the test was effectively worthless. The AAMC prepared to create new-question-only forms in case the courts reversed the injunction, but acknowledged doing so would increase the test's developmental expenses[93]. The science knowledge section was

---

attention it received; a memo from James B. Erdmann dated 10 June 1987 (AAMC Archives, "A.G. Swanson MCAT Papers" Folder 8) discusses a similar bill in less populous Wisconsin. Some action is recommended, but not to the level deployed to fight bills in New York and California.

[92] Most of these documents are not available, but a few were included as examples in the AAMC's 1988 Request for Proposals for a revised MCAT. See Sandra R. Wilson-Pessano, et al, *MCAT Verification Report, Spring 1987 Administration (Forms 38, 39, and 75)* (Palo Alto, Cal, AIR, 1987), p. 1 in AAMC Archives "MCAT: Essay" Box 501982739, Folder 29. Similar reports for the Fall 1987 and Spring 1988 administrations are included in Folders 31 and 32. The placement of unscored experimental questions is not explicitly mentioned, though such questions existed. The two year gap between reuse was intended to compensate for repeat examinees. It is unknown how other tests attempt to achieve a similar result, and whether they changed to comply with New York's ruling.

[93] Much of the correspondence found in Folder 6 of the AAMC Archives "A.G. Swanson MCAT Papers" discusses this possibility. Administrative expenses (which can be found in Folder 5 of the same box) suggest that the MCAT generated revenue only through extra score reports. The AAMC did generate

particularly threatened, since a relatively finite amount of questions existed for each assigned topic. Releasing these questions removed them from the potential question pool, increasing the need for new questions to be drawn while still remaining within the strict content outline. Perhaps more so than any other test, the MCAT relied on keeping test questions hidden from public view in order to preserve its effectiveness.

The measures taken to enhance the New MCAT required new levels of secrecy. Had New York taken aim on the old test, it is conceivable that the AAMC would not have offered such a fierce challenge. Though predictive validity was of major importance regardless of the test iteration, the New MCAT carried a set of limiting content restrictions threatened by New York's law. Pushing for added relevance and reliability bound the MCAT to reproduce the same information regardless of the date or content of a particular test. Though the situation dragged on throughout the eighties, a legal victory seemed the only way to avoid putting the MCAT into an even more precarious position.

TOWARDS A NEW "NEW MCAT"

The situation in New York set the tone for more conflict throughout the eighties. As applicant levels dropped, the quality of the remaining students became increasingly distressing. Premedical students were increasingly perceived as too competitive, obsessed with presenting the best possible medical school application at the expense of the things that did not fit on an AMCAS form. Only five years after introducing the New MCAT, the AAMC began exploring ideas to revise it; these early movements were the genesis of

---

substantial revenue through the AMCAS, however. It is unknown whether this remains the case, but it seems likely. The MCAT is currently $210, and the AMCAS costs $160 for the first school applied to and $30 for each additional school. Applying to thirty schools or more is not uncommon, and applying to one hundred schools is not unheard of.

another iteration of the MCAT first administered in 1991. This new version cut four science scores down to two, and returned to the idea of including a section intended to "motivate undergraduate students to strengthen their general education."[94] In light of these changes, the New MCAT seems like an overextension, an extreme response to a single crisis that ignored—or perhaps exacerbated—other problems facing medical education near the end of the twentieth century.

Even this extreme response, however, was heavily compromised. Based on the suggestions in the MCAAP regional conference *Position Papers*, one can see the origins of many of the New MCAT's features, but only amidst dozens of other suggestions that fell by the wayside. Neither the northeastern students proposing an achievement-only test nor the midwestern students in favor of aptitude-only testing could have been entirely satisfied to see both remain on the New MCAT. Though the MCAAP had been intended as an entire structural overhaul of the admissions process, little else changed beyond the standardized test, the one problem identified by a near-universal consensus. While there were many supporters of non-cognitive evaluations, personality metrics proved too difficult to accurately and efficiently standardize. Beyond attempting to synthesize so many different viewpoints, the New MCAT's designers had to preserve the test's reliability, or else risk the MCAT becoming irrelevant. Though the refined content was one of the AAMC's primary selling points, admissions committees did not need the MCAT to be more relevant to the curriculum, or even to be equivalent across ethnicities. As long as the numbers provided by the test offered a decent guess of student performance, medical schools seemed little concerned about what the MCAT specifically

---

[94] Robert I. Keimowitz, "Rx: Writing I" *AM* 64, no. 11 (Nov. 1989) p. 662. A longer discussion of this situation will be included in the second part of this research, "Rebuilding a Better MCAT: Premedical Syndrome, Standardized Essays, and the Medical College Admission Test, 1981-1993."

tested. The intentions of the MCAT had to work within the parameters of how the test was actually used, further adding to the compromise.

The difficulty of altering the MCAT—in light of the "somewhat bloody past history" of prior revision efforts—reveals an urgency behind the drastic revisions of 1977. Undoubtedly, educators had expressed concerns with the test's content well before applicant totals began surging forward, but overhauling the test did not yet seem like a dire necessity. As the weight of extra applications compromised the usefulness of admissions procedures, however, the need for change became clear. The MCAT was among the most glaring weaknesses in the process, and revising it was the simplest way to reach the widest number of people; thus, the AAMC used the New MCAT as its primary response to the changing face of medicine. Physicians needed to keep pace with the expanding scientific demands of medicine while practicing within the tightening bounds of legal liability and the standards and regulations that proliferated particularly in the post-Medicare period. Since a test given to undergraduates could not actively evaluate clinical acumen, the MCAT opted to search for success by locking its focus on the most essential scientific abilities required for success in these new conditions.

The ideal of the well-rounded physician did not die out during this period; the rare applicant showing excellence in all fields continued to find a place in medical school. But even though medical students represented some of the brightest students in any given year, most applicants were not quite so universally talented. Within this slightly lower tier, the New MCAT, in its design and particularly in how medical schools used it, favored those with a narrow scientific focus over those whose abilities in chemistry and biology were slightly lacking. This preference did not necessarily mean the latter group

of students could not become physicians, but rather, that the former group was simply a safer bet. Students likely to struggle in the first half of medical school could be, and often were, outstanding clinicians, but they could also fail to graduate. Those who demonstrated the ability to handle the first two years of medical school were almost guaranteed to finish the next two; some proved to be excellent doctors, while others did not, but almost none of them dropped out along the way. With such a surplus of qualified applicants, medical schools felt obligated to lower their already miniscule attrition rates; every failed student took a place that might have gone to a rejected applicant. There were too many unseen, unknown elements of admissions to know exactly which students most deserved acceptance, and the substantial public and institutional investments attached to it. Centering the MCAT on a stronger scientific foundation was not a perfect solution, but it was an effort to make sure the investments made in admissions would produce results.

For reference

Not to be taken from the room.