Questioning the Premedical Paradigm

Barr, Donald A.

Published by Johns Hopkins University Press

Barr, Donald A.
Questioning the Premedical Paradigm: Enhancing Diversity in the Medical Profession a Century after the Flexner Report.

For additional information about this book
https://muse.jhu.edu/book/462

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=8610
By 1920 premedical education had become largely standardized in the United States. Medical schools differed somewhat on the expected length of the premedical education—some expected only two years of college, while some required a bachelor’s degree. However, nearly all medical schools required college courses in chemistry, biology, and physics. Once a student had successfully passed these courses, he was then eligible for admission.

The creation of this national norm for premedical education was not based on scientific evidence linking it with a higher standard of professional practice for the new graduates. Rather, it was grounded in the widely held belief that medical education must by its very nature be based in science—both education in medical school and premedical education in colleges and universities. This was the model of medical education the United States had imported from Europe starting in the late 1800s.

As long as there was a place in medical school available to every undergraduate who had successfully completed the required sequence of premedical courses, the level of a student’s performance in those courses, either his absolute level or his level relative to his premedical peers, seemed less important. This was to change, however, beginning in mid-1920s, when for the first time the number of applicants to medical schools nationwide was greater than the number of the available places in medical schools.

Recall from the previous chapter that, largely as a result of the efforts of the American Medical Association (AMA), the Association of American Medical Colleges (AAMC), and a series of laws regulating medical practice passed in a number
of states, the number of medical schools in the United States decreased from more than 160 in 1910 (the year the Flexner Report was published) to 85 in 1920. Coincident with the rising professional status of the medical profession during this era, an increasing number of college students became interested in a medical career. Between 1926 and 1935 the number of applicants to medical school nationwide increased by 50 percent.\textsuperscript{1} Between 1926 and 1927 alone, applicants increased by 32 percent.\textsuperscript{2} The combination of a decrease in the number of medical school slots, accompanied by an increase in the number of medical school applicants, inevitably led to the need to develop mechanisms to select from among those applying for admission those students who were the most “fit to study medicine.”

The problems this situation presented were discussed in a paper presented to the annual meeting of the AAMC in 1926 by Dr. John Wyckoff of the New York University Medical College. In 1919 NYU had experienced for the first time a greater number of applicants than available slots. Beginning in 1921, NYU had created its first Admissions Committee, charged with selecting among these applicants. In describing the admissions criteria used by the committee, Wyckoff stated: “Obviously, three requirements are fundamental: mental equipment, physical equipment, and that quality so difficult to define—character. While it is undoubtedly true that a poor or mediocre student, if he has the usual character, will make a better physician than a man of high scholarship with less character, still, there is a minimum of mental ability that is essential if he is to carry the medical curriculum.”\textsuperscript{3}

By “carry the medical curriculum,” Dr. Wyckoff means not to fail the first years of medical school. Between 1910 and 1920, the number of NYU medical students who failed the first year of medical school ranged between 20 and 40 percent. Dr Wyckoff commented that “the usual wastage, which comes from a large percentages of failures at the end of the first and second year, is partly unnecessary and should be avoided.”\textsuperscript{4} To avoid this “wastage,” NYU began to look for an association between a student’s grades in the first year of medical school and his grades in the premedical sciences. The association was clear: the group of students with the highest premedical grades also had the highest medical school grades; those with the lowest premedical grades had the lowest medical school grades. Beginning in 1922, they used this association to select students for admission, accepting those students with the best grades in the premedical sciences. Within a few years the failure rate at the end of the freshman year had been cut to less than 6 percent. Commenting on the success of this new program, Dr. Wyckoff remarked, “It is interesting to see how the wastage at the end of the first year at medical school may be cut by giving heed to the collegiate standing of students.”\textsuperscript{5}

In 1928 Dr. Frederick van Beuren reported similar data from Columbia’s med-
ical school, coming to a somewhat different conclusion. While Wyckoff had looked only at undergraduate grades in the premedical sciences, van Beuren looked both at grades in the sciences and overall undergraduate grades. “We found, to our surprise, that the average grade of all the subjects studied was a better indication of character of the work of a student would do in the medical school than the average grade of the premedical required subjects alone.”

Before Wyckoff and van Beuren presented their data, there had been few published studies of factors that predict success in medical school. A literature review conducted by the American Council on Education and reported in 1929 of 3,650 reports of educational research published in the preceding ten years “found only seven related to medical education.” One of these seven, however, is of substantial importance and presents a somewhat different conclusion from Wyckoff’s study.

At the 1914 meeting of the AMA’s Council on Medical Education (CME), A. Lawrence Lowell, who had become president of Harvard University in 1909, cautioned delegates against adopting an overly rigid premedical education based primarily in the sciences, suggesting that such a curriculum “may have the effect of excluding able men from the profession.” Lowell cited research he had published in 1911 on the success of Harvard medical students. Looking at the undergraduate and medical school experiences of students who had attended both Harvard College and Harvard Medical School between 1895 and 1910, Lowell asked whether a student’s undergraduate performance in the sciences or his overall performance in his undergraduate studies was the better predictor of medical school success. Rather than using Wyckoff’s measure of success (first-year medical school grades), Lowell looked to see which students had graduated from medical school with cum laude honors distinction. The highest rate of honors distinction was among students who had focused their undergraduate studies in literature, languages, philosophy, or mathematics, leading Lowell to conclude “that natural science in college is certainly not a markedly better preparation for the study of medicine than other subjects.” Lowell acknowledged that “the young man who has acquired some familiarity with natural science and the use of instruments has, no doubt, an initial advantage in the study of medicine, and is much easier to teach at the outset” [i.e., the first year of medical school] . . . but that . . . initial advantage was soon overcome in the course of professional study.” Thus, as far as Lowell was concerned, “one subject is not distinctly better than another as a preparation for professional education.”

The quote at the beginning of this chapter by T. R. McConnell, founder of the Center for Studies in Higher Education at the University of California, Berkeley, sets out an important underlying issue in studies of the factors that predict or are associated with professional success. There are two ways of measuring success: aca-
demic success, and professional success. Grades in the first year of medical school represent a form of academic success; being recognized for distinction in overall medical school performance, pre-clinical as well as clinical, is a more general form of professional success. As we will see below, for much of the twentieth century, medical school admissions committees were concerned principally with predicting early academic success in medical school. It was only in the second half of the century that educators began to look more seriously at measures of professional success as indicated by the level of clinical skills of a practicing physician.

Predicting Early Academic Success in Medical School with Standardized Exams

By 1928 most medical schools were facing the problem of high rates of early failure among medical students. As described by Burton Myers, dean of the medical school at Indiana University, “The enrollment of 120 freshmen with the expectation of having 100 sophomores the following year, dropping 20 students whose year has cost an average of $700.00 per student, a loss of $14,000.00, is not economically justifiable if we can get our 100 sophomores by a more discriminating selection of 110 or fewer freshmen at a saving of $7,000.00 or more of school budget, the salary of a full-time staff man.” In speculating about how medical schools might avoid this “wastage,” as Wyckoff had referred to it, Myers cited a study from the educational psychology literature published in 1923 in which Mark May of Syracuse University had used the results of two separate intelligence tests given to 450 incoming college students to predict their grades in college. Myers concluded that “the most reliable means of predicting academic success is a combination of intelligence and degree of application [i.e., effort].”

In 1923 the use of intelligence tests was fairly new. One of the first was developed in 1905 by French psychologist Alfred Binet and was used to identify young children who were likely to have trouble in school due to their sub-par intelligence. Binet was later to collaborate with Lewis Terman at Stanford University to define the concept of intelligence quotient, or IQ, as the ratio of a subject’s mental age (as measured on the new intelligence test) to the subject’s physical age. As described by Nicholas Leman, in the years leading up to World War I, Terman and others “were tireless advocates of the widest possible use of IQ testing by American educators, so that students could be assessed, sorted, and taught in accordance with their capabilities.”

In the early use of the concept of IQ, intelligence was seen as an inherent human trait, something with which one is born. Intelligence tests could be used to identify slow learners who needed extra help in school, and it could be used to identify those
with the potential for higher education. “The idea of IQ testers was not to reform education, especially higher education, so much as to reserve it for highly intelligent people, as indicated by IQ scores, lest their talents be wasted.”

When World War I came in 1914, the U.S. Army arranged for Prof. Robert Yerkes of Harvard to administer an IQ test to nearly two million recently recruited soldiers in order to identify those recruits best suited for training as officers. The success of IQ testing in this regard substantially increased both the general awareness of and the belief in intelligence testing as a valuable educational tool.

Evaluating the intelligence of physicians in the army during World War I yielded some interesting results. It turns out that Medical Officers in the army scored consistently lower on the battery of intelligence tests than did officers in the Engineer or Field Artillery Corps. Further sub-set analysis showed the measured intelligence of Medical Officers varied substantially according to the AMA classification of the medical school from which they graduated: those graduating from schools ranked as “Class A” scored the highest on the army’s intelligence test, while those graduating from schools ranked as “Class C” scored the lowest. It is also interesting to note that graduates of homeopathic schools of medicine scored substantially higher than even the graduates of “Class A” “regular” schools.

Carl Campbell Brigham was a psychology professor at Princeton. Working to adapt the intelligence test used by the army during World War I for use in a broader educational context, Brigham used a combination of mathematical calculations, identification of facial expressions, and word recognition to create a new test to use in the assessment of the intelligence of would-be college students: the Scholastic Aptitude Test, or SAT. The SAT, later to become the national standard in assessing the academic qualifications of high school students, was administered for the first time in 1926 to 8,040 high school students who were applying to college.

The American Council on Education (ACE) is an organization that represents colleges and universities. In 1929 its assistant director, David Allan Robertson, addressed the AMA’s Annual Congress on Medical Education. Robertson described the work of Dr. F.A. Moss of George Washington University Medical School in adapting the SAT for use in evaluating applicants to medical school. Moss’s “scholastic aptitude test for medical school” was both a test of general intelligence and a test of one’s knowledge of the premedical sciences. It included six sections:

1. a test of scientific vocabulary
2. a test of premedical information
3. a visual memory test based on having viewed for ten minutes a diagram of the heart and the major blood vessels
4. a verbal memory test based on having read a paragraph about the heart and the major blood vessels
5. a reading comprehension test
6. a test described as “understanding of printed material”

Moss had administered this test to the 1927 freshman medical school class at George Washington and, based on the test results, predicted which students would fail in medical school and which would attain academic distinction. Eight of the ten students predicted to fail did so; six of the eight students predicted to attain distinction did so. Based on these results, the ACE printed large quantities of the new test and handed them out to delegates to the 1929 meeting to be used in assessing their current first-year students. Those who administered the tests could send them to Dr. Moss for scoring. Robertson noted that “Dr. Moss has undertaken to send the results to the deans in time for them to use the scores, if they so desire, in connection with the elimination of students at the close of the present year. Obviously, if convenient tests which will reliably predict academic success in professional schools can be worked out, a great waste can be avoided for the individuals and institutions which now are losing time and energy in trying to make educational adjustments which cannot be made.”

From the outset, the scholastic aptitude test for medical schools, later to become the MCAT, was used principally to weed out applicants who were predicted to fail the first year of medical school. By using the test to define and measure the level of scientific knowledge required to predict success in the first two years of medical school, Robertson suggested, it was then the job of the undergraduate institution, “to provide a curriculum more directly effective in training men and women for the medical profession and in helping to choose them wisely.”

Twenty-six medical schools administered the “medical aptitude test” (MAT) developed by Moss to their freshman medical school class, forwarding the tests to Moss for scoring and providing Moss with the students’ first-year grades. Moss divided the approximately 900 students into deciles based on their test scores, and then sorted students grades into four categories: 90 or greater, 85–89, 75–79, less than 75 (described as “failure”). Presenting his results to the annual meeting of the AAMC held in 1929, he reported a clear association between MAT scores and first-year grades. Of the students in the top decile, none failed, and 93 percent had grades of 80 or higher; of students in the bottom decile, 42 percent failed, and only 14 percent had grades of 80 or higher. The overall correlation between MAT score and grades was 0.59. This compared to a correlation between undergraduate grades in the premedical sciences and first-year medical school grades of 0.50.
While his results were impressive, Moss pointed out to the delegates a potential problem. While 42 percent of students in the bottom decile failed the first year of medical school, 58 percent passed all their courses, albeit with lower grades than many of their classmates. If, in an attempt to prevent the future failures from entering medical school, admissions committees had administered the MAT to these students as applicants and refused admission to all students scoring in the bottom 10 percent, a substantial number of students fully capable of passing the medical school curriculum would have been refused admission as well.

Moss developed what he referred to as a measure of the “efficiency” of an admissions screening criterion by comparing the percent of failures that would have been prevented by using a criterion to screen out applicants with the number of students in his sample attaining a grade of 85 or higher in their first year of medical school who would have been refused admission based on the device. He noted, “We secured the best results by combining the Aptitude Test scores with the premedical grades. When such a combined criterion was applied to the group on which records were available, we found that 94 percent of the failures would be eliminated, and 20 percent of those who would make 85 or above. . . . It is quite probable that the ideal method for selecting students will be a combination of this method with the results of the aptitude tests and the premedical grades.”

Moss proposed to the meeting that all schools in the AAMC begin to administer the MAT to applicants for admission and that they do it nationally on the same day. His office would take responsibility for scoring the tests and reporting the scores to the deans. Two motions were made to the delegates at the meeting: (1) “that the Association record its sense of the importance of the study of aptitude tests in relation to the acceptance of students in medical schools”; and (2) that “the Association appoint a committee to direct an experimental study of aptitude tests for admission to medical studies” in the manner suggested by Moss. Both motions passed, apparently enthusiastically.

The newly established Special Committee on the Evaluation of the Aptitude Test for Medical Students took on the study of the MAT and how it should be used in the admissions process, reporting back to the AAMC on a regular basis. In 1935 the Special Committee reported on its research to date. Responding to Moss’s concern that use of the test to eliminate potential failing students would also eliminate a substantial number of students who would pass their medical school courses, the committee suggested that “a common sense practical view of the problem would seem to be one in which it is admitted that the best criterion is the one which would eliminate the greatest number of failures and at the same time the fewest number of good students. . . . In a very real sense there is and can be, probably, no right or cor-
rect answer to the problem.” At a time when 20 percent of entering medical students had failed by the end of their first year, it is understandable that the AAMC and the deans of the medical schools were willing to refuse admission to otherwise capable students based on a low test score in order to reduce the number of failing students.

Not all those who read the committee’s report agreed with this approach, however. Edward Thorndike, a leading educational psychologist from Teachers College at Columbia University, wrote the following comments: “Superficially, the tests look somewhat pedantic and over-specialized and over-weighted with memorizing; and they probably are better to predict success in the first two years of medical school than success later and throughout life. I imagine they are frankly designed to weed out the kind of persons who would be weeded out by the first two years of work in medical school.”

The committee reported additional data in 1938 and in 1940. For the entering medical school class of 1936–37, 84 percent of all entering students had taken the MAT. Of students scoring in the lowest decile of test scores, 25 percent failed the first year of medical school (and of course, 75 percent passed the first year). Only 2 percent of students in the highest decile failed in the first year. In 1938–39, with 90 percent of entering medical students having taken the test, the failure rates were nearly identical: 22 percent of students in the bottom decile, and 3 percent in the top decile. Moss felt that he had convincing evidence that the MAT score, taken together with a student’s grades in premedical sciences, provided the best tool for predicting which students would fail the first year of medical school.

Moss paid less attention to what happened to students after their first year or two of medical school. The issue of the clinical or professional skills ultimately developed by students seemed of little concern to him. In 1933 W. F. Kramer of the University of Chicago pointed out that “success in medical schools is best measured by the success of the graduates after they leave school.” This view was echoed by I. L. Kandel, professor of education at Teachers College of Columbia University. In a report commissioned by the Carnegie Foundation for the Advancement of Teaching (the original publisher of the Flexner Report), Kandel reviewed the use of aptitude tests in the admissions process of schools of medicine, law, and engineering, concluding that “aptitude tests can only discover whether a candidate is likely to succeed in the professional preparation selected. They do not indicate promise of future success in the practice of that profession.”

In the 1940 Special Committee report, Moss made an important observation: “We found that questions taken directly from the premedical sciences have a much higher selective value than do general cultural questions based on knowledge of art,
music, drama, history, literature, etc., or questions based on geography and current events. . . . As a result of this study we have greatly increased the number of pre-medical information questions and practically eliminated questions of a more general type in constructing the new form of the test.”

The MAT was becoming less a test of general scholastic aptitude and more a test of familiarity with the premedical sciences.

World War II and its aftermath brought a substantial increase in the number of students applying to medical school. Both the need to train doctors for the war and the entry of returning veterans into the educational system added even more pressure to those evaluating applicants for admission to medical school. Officials at the AAMC thought that Moss’s MAT continued to have shortcomings in its ability to select among applicants most efficiently, so in 1946 they replaced it with a new test called the Professional Aptitude Test. In 1948 this test was renamed the Medical College Admission Test (MCAT), the name it has today. In a comparison of the MAT and the MCAT, R. B. Ralph and C. W. Taylor emphasized that, in the face of the rising number of applicants, “the task of selecting those best fitted for medical training and of eliminating misfits at the earliest possible moment becomes increasingly important.”

Unfortunately, in comparing the power of the MCAT to that of the older MAT to predict grades in the first two years of medical school, Ralph and Taylor concluded that various parts of the new test “have zero or negligible value as predictors.”

Throughout the 1950s researchers continued to try to improve the process by which students were selected for medical school. By 1959 the MCAT had been modified and had four sub-sections: Verbal, Quantitative, Modern Society, and Science. A separate score was reported for each section. In a study of more than 12,000 students applying to the State University of New York College of Medicine in Brooklyn between 1950 and 1957, J. K. Hill found that the combined score of the science and quantitative sections had the strongest association with academic success in medical school, again measured as grades in the first year of school. The association between the Verbal Ability score on the MCAT and freshman success was substantially lower.

Not everyone associated with medical school admissions was comfortable with the continuing emphasis on predicting success and avoiding failure in the first year of medical school. In a 1957 review of research on medical school admissions, Gottheil and Michael cautioned, “Presumably, the goal of medical education is to produce ‘good’ doctors of medicine. What constitutes the good doctor however, and how to evaluate the constituent factors remains the most perplexing problem in the field. . . . The use of medical school grades as a criterion against which to evaluate
the success of a selection program is not only subject to criticism on the grounds that grades may not be correlated with the quality of later practice of medicine, but there is an even more basic idea to consider: whether medical school grades are in themselves statistically reliable.” The authors went on to ask, “To what extent can or should a broad cultural background in the socio-humanistic field be sacrificed for outstanding achievement in science?”

**Broadening the Scope of the Admissions Assessment to Include Predictors of Clinical Performance**

By the 1950s, a number of the leaders in medical education in the United States were becoming concerned with the overemphasis on using success in the premedical sciences to select students for medical school. To address this issue, the AAMC convened a four-day teaching institute in 1956 at which representatives from most U.S. medical schools met to discuss “The Appraisal of Applicants to Medical School.” The conference was to address the following question: “Is medicine attracting those students who are best endowed with the characteristics most favorable for serving the health needs of society and the research needs of medical science?”

In addressing this question, the AAMC first administered a survey to administrators and admissions committee members at 91 medical schools in the United States and Canada. They then held a series of panel discussions and workshops to discuss the results of the survey. As reported by Dr. Robert Glaser, dean of the University of Colorado School of Medicine, the survey largely confirmed the heavy historic emphasis placed on performance in the premedical sciences, and on MCAT scores as a reflection of that performance. Eighty-six percent of the schools reported placing great importance on science grades in evaluating applicants for admission, while 40 percent reported placing great emphasis on non-science grades. Fifty per cent of schools also placed great emphasis on MCAT scores. Among the premedical sciences, schools reported placing most emphasis on grades in the natural sciences, especially chemistry and physics, and only to a lesser degree biology. Similarly, schools reported giving most emphasis to an applicant’s MCAT Science and Quantitative scores and relatively little emphasis to their Verbal Ability or Modern Society scores, leading Glaser to comment that “knowledge of modern society as measured by MCAT is not considered to be of major importance in the evaluation of the applicant’s intellect.”

It was in response to statistics such as these that T. R. McConnell of the Center for Studies in Higher Education at the University of California, Berkeley, made his remark that leads off this chapter. What are we trying to do? McConnell asked.
we trying to select those students who will do well academically in the early part of medical school, or are we trying to select those students who will make the best physicians after medical school? The two outcomes are not necessarily the same. This issue received substantial attention during the conference. Dael Wolfe, the executive officer of the American Association for the Advancement of Science, echoed McConnell’s remarks. In trying to select the most qualified students, he said, “we must face the problem of deciding more highly qualified for what? More highly qualified in terms of what measures?” R. F. Arragon, a professor of history from Reed College, concurred in the need to look beyond early success in the sciences, commenting that “there does seem to be some general assumption that there are qualities that may be necessary for success in the first two years—different qualities from those necessary for the clinical years to follow.” Commenting on the need to look beyond early medical school grades, Robert Glaser suggested, “Perhaps it is overly optimistic to suggest at this time that sound means of evaluating actual physician practice can be developed and that eventually selection measures may be validated against these more ‘ultimate’ criteria.”

The discussions at the conference of the need to broaden the perspective used in evaluating applicants to medical school were summarized by John Caughey, then associate dean at the Western Reserve School of Medicine:

The principal result of the discussion of this topic was the realization by the participants of the great need for continuing well-organized study of medical student selection. . . . However the real challenge lies ahead and has not been accepted by medical faculties and admissions committees. This challenge is to define more precisely the expectations we have for members of the medical profession, to determine the intellectual and personal qualities which are necessary for the roles they are expected to play, and then to find means to attract, select, and educate the kind of students who, as physicians, will strive with reasonable hope of success to make the desired contributions to medical education, scientific research, and the health needs of their community.

The AAMC conferees went on to discuss at some length the importance of including assessment of the nonintellectual characteristics of applicants as well as measures of their intellectual achievements, an issue I address in the chapter that follows.

The conference’s emphasis on the pressing need to find ways to look beyond the first two years of medical school in gauging medical student success soon began to be reflected in the literature on premedical education. An important series of papers responding to this need began to appear in the early 1960s. In 1962 Schwartzman and colleagues from McGill University in Canada reported on their study of the as-
sociation between the traditional markers of undergraduate performance and medical student grades in each of the four years of medical school. In looking at performance beyond the first year, they identified several important relationships:

- While there was an association between MCAT scores and student performance across all four years of medical school, the relationship was not as strong as had been previously reported of studies looking only at performance in the first year.
- There was an association between grades in the five required premedical subjects (the four sciences plus English) and student performance in the first year of medical school.
- By the fourth year there were no significant relationships between premedical grades and performance, although organic chemistry grades and English grades showed a weak association.

In 1962 Funkenstein looked at which students leave before completing medical school. Rather than looking principally at who leaves after the first year, as most previous studies had done, he looked at students who left for any reason across all four years of medical school. He confirmed that the highest dropout rate was after the first year, with 5.5% of students leaving. The dropout rate decreased significantly after that: 2.1% after the second year, 1.1% after the third year, and 0.3% during the fourth year. Once students made it through the first year, nearly all were successful in completing medical school. A later study by Gough and colleagues confirmed the substantially lower dropout rate after the first year, and indicated that those students who dropped out of medical school during the clinical years did so largely for personal rather than academic reasons.

However, Funkenstein did notice a distinct pattern: those who dropped out during the first two years tended to be weaker in their premedical science and stronger in premedical humanities, while those who dropped out during the final two clinical years tended to be stronger in their premedical science and weaker in premedical humanities. A series of papers by Korman and colleagues supported the concept that, during medical school, students who were stronger in the undergraduate sciences than in the humanities tended to have different experiences and pursue different career goals than their colleagues who were stronger in the humanities.

Richards and colleagues looked beyond medical school to assess the associations among premedical grades, MCAT scores, grades in medical school, and performance in the internship year immediately following medical school. The internship assessments reflected a global evaluation from the internship director of the intern’s clinical skills. The authors concluded that “the best predictor of intern performance
is grade average in the clinical year(s) of medical school, and that grades in the preclinical years of medical school [i.e., the first two years] have only a slight relationship to intern performance, and that premedical grades have almost no relationship.”

Interestingly, the authors noted a negative but non-significant association between MCAT scores and intern performance, raising the possibility that the better a student did on the MCAT, the less well he or she did as an intern. Howell and Vincent also found a negative association between MCAT scores and evaluations of the clinical quality of interns.

Johnson and colleagues took their evaluation one step further, looking at the association between medical school performance and clinical performance in a multiyear residency. They did not break down their assessment of medical school performance by year of school, but rather looked at a student’s relative class standing across all four years. While students who ranked higher during medical school tended also to rank higher as residents, there was substantial crossover, with a number of lower-ranking medical students becoming high-ranking residents, and vice versa.

Price and colleagues went beyond evaluations of postgraduate medical training to look at the professional skills of a sample of about 500 practicing physicians, representing academic practice, urban specialty practice, and both urban and rural general practice. They calculated a composite score of professional quality from a range of individual measures and then compared this score with premedical grades and medical school grades. The authors concluded, “Our study clearly demonstrates that performance in formal education, as measured by grade-point averages, comes out as a factor almost completely independent of all the factors having to do with performance as a physician.”

The *Journal of Medical Education* in which the Price and colleagues paper was published was the official journal of the AAMC. Following the Price article, the journal published the transcript of a discussion of Price’s presentation of his results that had taken place at an AAMC meeting. That discussion posed a very interesting question, one that has continued relevance today. In response to the paper, Dr. George Saslow of the University of Oregon asked, “Suppose one of us had the power to start off a new medical school with a faculty willing to listen to data like this. In what directions would you suggest that we look in order to make predictions about the kinds of doctors that we need?” In response, Dr. Price replied: “The impression has grown on me more and more that since conventional grades and other measures used have been overweighted, difficult as it is, we are going to be forced to pay more attention to other qualities of character and personality, of behavior, of relationships to people, of matters of dedication and integrity. These things are hard to define and
difficult to measure, but they may be the most important factors, and it may well be that they can be determined to some extent in medical students.”

Between 1963 and 1973, three separate groups of authors published comprehensive reviews of the literature linking academic performance and subsequent clinical skills. Each supported the conclusion that the association between premedical performance and early medical school performance on the one hand with eventual clinical quality on the other was tenuous at best. Regarding faculty assessments of clinical quality in the fourth year of medical school, Gough and colleagues went so far as to suggest that “the MCAT scales and the three indices of premedical scholastic performance show an essentially zero relationship with this criterion.” Win- 
gard and Williamson also found, “little or no correlation” between premedical grades and clinical performance.

By the 1970s, medical schools had been using a combination premedical science grades and MCAT scores for more than forty years to select those students who were the most “fit to study medicine,” as originally described by Daniel Coit Gilman in 1878. What if the criteria they had been using were not optimal? What if we could improve the overall clinical quality of the medical profession by using a different set of criteria to select from among the many applicants to medical school? If we could start from scratch, and given the growing body of research on the predictors of professional success in medicine, how would we structure our admissions process? Of course, we can’t ignore history, nor can we expect members of medical school admissions committees simply to abandon processes that have evolved over a period of decades. However, the exchange between Drs. Saslow and Price raises intriguing questions.

**Broadening the Effort to Predict Clinical Quality in the Selection of Medical Students**

Research appearing in the 1980s and beyond and looking at factors associated with success in medical school typically included measures of clinical quality as well as academic quality. Clinical quality was often measured as performance in the clinical clerkships in the final two years of medical school and in the first postgraduate year of clinical training. In one such study, DeVaul and colleagues took advantage of a natural experiment in which a public medical school was instructed by its state legislature to expand its entering class after the notices of acceptance and rejection had already been sent out. This unexpected expansion of medical school slots permitted the admissions office at the school to compare the medical school success of 50 students initially rejected but subsequently accepted with that of 150 students initially
accepted. The authors concluded, “In attrition and in both pre-clinical and clinical performance through medical school and one year of postgraduate training, there were no meaningful differences between the groups.”

While there was general consensus on the need to include assessments of both academic performance and clinical quality, there was some concern that the measures used to assess clinical quality—typically the qualitative assessment of a clerkship director or internship director—did not provide as reliable a measure as did grades or standardized tests. Accordingly, researchers began to use a second measure of clinical quality in their research: scores on the national licensure examination administered by the National Board of Medical Examiners (NBME). Founded in 1915 as an independent nonprofit organization, NBME was charged with developing and administering a standardized licensure examination nationally. All medical graduates who wish to obtain a license to practice medicine must pass this exam. The exam was given in three parts: NBME I, testing knowledge of the preclinical sciences, given at the end of the second year of medical school; NBME II, testing clinical knowledge, given at the end of the fourth year of medical school; and NBME III, testing the application of clinical skills, given at the end of the first year of postgraduate studies (internship or residency). Using scores from these standardized examinations, researchers were able to have a more complete measure of success in medical school, to which they could compare measures of success in premedical studies, as illustrated in figure 4.1 below.

Using this general model of measuring outcomes of premedical and medical education, researchers were able to gain a more complete picture of those factors that predict success in medical school at the various stages of medical training. For example, in 1990 Dr. Karen Mitchell, vice-president for research at the AAMC and the director of the MCAT program, published her review of the literature linking premedical performance and medical school performance using this model. Using MCAT scores beginning in 1977, when the MCAT was reformulated to include more questions relating to scientific principles while eliminating questions pertaining to general knowledge of the liberal arts, Mitchell found that a combined measure of undergraduate GPA and MCAT scores was highly correlated with grades in the pre-clinical sciences ($r = 0.49$). Undergraduate performance had a weaker correlation with grades in the clinical years ($r = 0.38$) and with subjective assessments in the clinical years ($r = 0.27$). Similarly, undergraduate performance had the strongest correlation with the NBME I score ($r = 0.58$), less with the NBME II score ($r = 0.49$), and the weakest correlation with the NBME III score ($r = 0.35$).

In 1993 Glaser and colleagues published their study addressing the following question: Among science, verbal, or quantitative skills, which is the best predictor
of physician competence? In a sample of 1628 graduates of Jefferson Medical College who had entered between 1978 and 1985, they used three MCAT scores as indicative of undergraduate performance: science problems, reading skills, and quantitative skills. They compared these measures with success in medical school as measured by parts I, II, and III of NBME and found that:

- Scores on the science problems subtest were better predictors of the basic science component of physician education (NBME I scores) than were the reading scores.
- Both science problems and reading skills predicted clinical science scores equally well (NBME II scores).
- Reading skills scores contributed more than the science problems subtest in predicting scores on an examination of patient management skills (NBME III scores).
- Scores on the quantitative skills subtest did not contribute to any prediction.\(^{56}\)

From these results the authors concluded “that the verbal ability reflected in the reading skills scores of an applicant to medical school are more important indicators of later physician competence (as measured by standardized certifying examinations) than the applicant’s ability to solve scientific problems.”\(^{57}\)

In a study of two graduating classes from a single medical school, Loftus and colleagues looked at predictors of performance in the first year of residency. They
found that (1) subjective assessments of a student’s performance in the clinical clerkships in medical school were the best predictor of performance in residency, and (2) undergraduate grades (science and non-science combined) had little relevance to performance in residency.\(^{58}\)

Once researchers began to take a longer-term view of success in medical school, some clear patterns of associations began to emerge. In the era of approximately 1930–57, when researchers were concerned principally with preventing failure in the first year of medical school, it seemed both adequate and appropriate to use a student’s performance in the premedical sciences, measured either as grades or science MCAT scores, to predict medical school success. However, when researchers began to expand their view of medical school success, they found that the factors that predicted success in the first two years of medical school were only weak predictors of success in the final two years of medical school or of success in postgraduate training (internship or residency). The factors that were linked most strongly with these later measures of professional success were skills in the humanities and general verbal skills.

In an effort to improve its predictive ability, the MCAT was revised in 1977 and again in 1991.\(^{59}\) In order to balance the assessment of verbal ability and scientific knowledge, the 1991 version was divided into four parts: Biological Sciences, Physical Sciences, Verbal Reasoning, and Writing Sample. Descriptions of the specific content areas for the four tests are available on the Web site of the AAMC.\(^{60}\)

In 1996 research staff from both the AAMC and the NBME began to publish their studies of the new format for the MCAT and its ability to predict success in medical school. In one of the first studies, Swanson and colleagues looked specifically at the accuracy of the new test in predicting the first step of the national licensing exam. (In 1992 the format of the NBME examination was changed somewhat, and the name of the exam was changed to the United States Medical Licensing Examination [USMLE]; it was still developed and administered by the NBME and was administered in the same three steps as the previous NBME exam.) The authors described the purpose of the new MCAT as “to encourage students interested in medicine to pursue broad undergraduate study in the humanities and social sciences, as well as biology and the natural sciences. It emphasizes mastery of basic biology, chemistry, and physics concepts; facility with scientific problem solving and critical thinking; and writing skills.” In a study of 11,145 medical students, they found that the biological sciences and physical sciences components of new test were accurate predictors of USMLE I scores; that the verbal reasoning and writing sample scores had little predictive ability of USMLE I scores; and that after taking into account scores on the biological sciences and physical sciences components of the
MCAT, neither undergraduate science grades nor undergraduate non-science grades added to the predictive accuracy of the MCAT scores alone.61

A study by Wiley and Koenig looked more carefully at the issue of the added value of premedical grades after taking into account MCAT scores. Taken alone, the correlation between grades and USMLE I scores ($r = 0.43$) was not as strong as the correlation between MCAT scores and USMLE I ($r = 0.72$). However, when grades and MCAT scores were taken together in a test of multiple correlations, the combination of the two measures added little to the association with the USMLE I scores (combined $r = 0.75$). When the authors looked at the association between premedical grades and MCAT scores with grades in the first two years of medical school, their results were essentially the same. This paper confirmed that MCAT scores alone are essentially as good at predicting USMLE I scores or grades in the first two year of medical school as is a combination of MCAT scores plus premedical grades.62

Dr. Ellen Julian, director of the MCAT for the AAMC, reported a follow-up study that she described as “a comprehensive summary of the relationships between [undergraduate] GPAs and MCAT scores and (1) medical school grades, (2) USMLE Step scores, and (3) academic distinction or difficulty.” She noted a general pattern of decreasing rates of academic difficulty and increasing rates of academic distinction as MCAT scores increase. However she cautioned, “that incidents of distinction occur for students with very low MCAT scores, and incidents of difficulty occur for students with very high MCAT scores.” Regarding the issue of the relative effects of MCAT scores and undergraduate grades (uGPAs) in predicting all aspects of medical school success, she concluded, “MCAT scores almost double the proportion of variance in medical school grades explained by uGPAs, and essentially replace the need for uGPAs in their impressive prediction of [USMLE] Step scores.”63

**Separating the MCAT into Its Constituent Parts**

If MCAT scores seem to be the best predictor of success in medical school, the next question to address is whether the various sections of the test (biological sciences, physical sciences, verbal reasoning, writing sample) have similar or different associations with various levels of medical school success as measured by the three steps of the USMLE (scientific knowledge, clinical knowledge, clinical skills). Two recent research reports examined this question.

Veloski and colleagues looked at the records of several hundred medical students who entered Jefferson Medical College in the 1990s.64 As measures of premedical
preparation they looked at students’ undergraduate GPAs in their science courses and their MCAT scores. For the MCAT scores they included the verbal reasoning score and an average of the physical sciences and biological sciences scores. Using multivariate analysis to take into account students’ age, gender, and race/ethnicity, the researchers looked at the correlations between these measures of premedical attainment and each of the three steps of the USMLE. The multiple correlation coefficients for these analyses are shown in figure 4.2.

From the results of these analyses we see three patterns:

1. Both the MCAT science scores and the undergraduate science GPA have the strongest correlation with the USMLE I score (scientific knowledge), less with USMLE II (clinical knowledge), and least with USMLE III (clinical skills).
2. For steps I and II of the USMLE, the MCAT science scores have a stronger correlation than does the undergraduate science GPA.
3. The MCAT verbal score has the weakest correlation with the USMLE I score, more with USMLE II, and most with USMLE III (clinical skills). Among the three measures, the MCAT verbal is the strongest predictor of USMLE III.

Figure 4.2. Predictive validity coefficients (r) of various measures on USMLE Step Scores
Source: Data from Veloski et al., 2000
Based on Veloski’s research, it appears that it would be optimal to give the MCAT science scores and the MCAT verbal scores approximately equal weight in the evaluation of applicants to medical school because the combination of the two will give the strongest prediction of success throughout the various stages of medical education.

Donnon and colleagues reported a similar analysis of the association between MCAT scores and USMLE scores. They were able to undertake a meta-analysis of results from 23 separate studies reported between 1991 and 2006, involving more than 27,000 medical students. Their individual analyses had sample sizes ranging from 650 for testing the association between individual MCAT test scores with USMLE Step III scores and 15,000 for testing the association between individual MCAT test scores with USMLE Step I scores. They did not include undergraduate GPAs in their analysis. In addition, few of the studies they included in their analysis had data about subject age, gender, or race/ethnicity, so they did not include these variables. Their results are shown in figure 4.3.

The results of Donnon’s study are consistent with those of Veloski’s study, with a few interesting differences. The predictive power of the science MCAT scores is once again strongest for USMLE I and weakest for USMLE III. The physical sci-
ences MCAT score showed no association with USMLE III. The strongest predictor of USMLE III is once again the MCAT verbal score. In these analyses the MCAT writing sample, added to the MCAT in 1991, had little if any association with performance at any level of medical education.

An earlier report by Hojat and colleagues suggested that, while the writing sample had no association with MCAT science scores or USMLE I scores, better scores on the writing sample were associated with higher scores on the MCAT verbal and on USMLE II. In addition, the authors found that those who did better on the writing sample had higher non-science GPAs as undergraduates but similar science GPAs. The authors were also able to obtain results of a previously validated assessment of clinical skills displayed in the first year of residency, as completed by the residency director. They found that students who did best on the writing sample also scored higher in three areas of clinical skills: data-gathering and processing skills, socioeconomic aspects of patient care, and physician as patient educator. From these analyses the authors reported:

The findings of the present study confirm the research hypothesis that scores on the Writing Section of the MCAT yield a closer association with measures of clinical competence than with achievement in the basic sciences. . . . Therefore, it can be concluded that the Writing Sample measures a unique skill, different from those measured by the other sections of the MCAT, including the Verbal Reasoning section. It can be speculated that such a unique skill might be attributed more to factors that are not associated with achievement in sciences. Such speculation needs to be verified further by empirical evidence.

Evaluating Medical Students’ Performance in an Actual Clinical Setting: The Standardized Patient Examination

In 2004 a new component was added to the USMLE Step II examination: the Standardized Patient Examination (SPE). Similar to the clinical skills assessment reported above by Hojat, this examination was intended to measure the clinical skills of students in their fourth year of medical school by observing them in a series of encounters with patients. In order to standardize the evaluation of students’ clinical skills, the NBME hired and trained laypeople to act as patients in order to be able to give a consistent history suggesting a specific medical problem and in some cases to mimic certain physical findings. Over the period of one day, a student evaluates twelve different patients. Each standardized patient is visited by a series of students. Students are scored on a pass/fail basis based on evaluations by the standard-
ized patient and by a trained physician-evaluator. Students must pass this examination in order to be eligible for licensure.

Several individual medical schools have been using standardized patient evaluations for some time as part of the assessment of medical students’ clinical skills. In 1992 Vu and colleagues reported on the use of SPEs at the Southern Illinois School of Medicine. Comparing SPE scores with scores on the NBME I and NBME II, they concluded that “the three types of measures did not rank the students similarly and may not have assessed all the same skills.” The suggested that faculty use a combination of the three types of examinations to evaluate students. Colliver and colleagues reported that the standardized patient’s satisfaction with a student’s interpersonal and communication skills during the exam was closely related to the student’s skills in history taking and physical examination. Basco and colleagues found little association between the SPE scores of third-year students and those students’ undergraduate GPA or MCAT scores. Similarly, Edelstein and colleagues found little or no correlation between SPE scores and either undergraduate GPA or MCAT scores; they did find moderate correlations with USMLE Step I (r = 0.25) and Step II (r = 0.30).

While the United States adopted the SPE in 2004 as part of the USMLE sequence, the Medical Council of Canada (analogous to the NBME in the U.S.) added an “objective structured clinical examination” (OSCE) to their licensing examination in 1992. The OSCE involves brief encounters with twenty standardized patients and is scored on a numeric basis. In addition to the OSCE, the Canadian licensing examinations include a “Declarative Knowledge” section (MCC Part 1) and a “Clinical Reasoning Skills” section (MCC Part 2). The MCC Parts 1 and 2 are quite similar to the USMLE Steps I and II. Similar to studies in the United States,

- the MCAT Biological Sciences score is correlated with the MCC Part 1 score (r = 0.19) but substantially less so with the Part 2 score (r = 0.03)
- the MCAT Verbal Reasoning score is correlated both with the MCC Part 1 score (r = 0.26) and with the Part 2 score (r = 0.24)
- the MCAT Physical Sciences score is correlated neither with the MCC Part 1 score (r = −0.03) nor the Part 2 score (r = 0.02).

In 2007 Tamblyn and colleagues reported a long-term follow-up of 3,424 physicians in Canada who had taken the OSCE between 1993 and 1996. Comparing the physicians’ scores on the OSCE with their MCC scores, they found the correlations shown in table 4.1.

As one might expect, a physician’s ability to communicate well with patients is more strongly correlated with the MCC Part 2 than with the MCC Part 1. The data
acquisition and problem-solving sections of the OSCE have the reverse pattern of
correlation. The magnitude of the correlation between the communication score
and the MCC scores was lower than that of the other parts of the OSCE.

The authors then looked at the frequency of quality of care complaints filed with
regulatory authorities against the physicians in the study. They found that “lower
[O]CSE communication scores were associated with a higher rate of retained com-
plaints, particularly in the lowest quartile of these scores.”76 In an editorial accom-
panying the Tamblyn article, Makoul and Curry responded to these results by rec-
ommending that, in order to improve quality of care, “initiatives could include more
systematically assessing interpersonal skills during the admissions process . . . and en-
suring that clinical skills assessments include a communications component.”77

The SPE and the OSCE appear to offer a valuable additional means of assessing
the clinical and professional skills of medical students at the completion of their
medical education. As the traditional measures used to evaluate students for ad-
mission—science GPA and MCAT scores—appear to have little power to predict the
clinical skills measured by the SPE, we will need further research to identify which
characteristics of applicants, both cognitive and noncognitive, provide the best pre-
diction of a student’s future level of these clinical skills.

### A Need to Rethink the Criteria We Use
to Select Medical Students

When in the 1920s medical schools first started using measures of premedical per-
formance to sort and select students for admission, the principal concern for ad-
missions officials was to reduce the number of admitted students who failed the
medical curriculum. That curriculum had only recently become grounded in sci-
ence, with nearly all medical schools adopting the four-year, science-based curricu-

---

**Table 4.1.**

<table>
<thead>
<tr>
<th>OSCE Section</th>
<th>MCC Part 1</th>
<th>MCC Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient communication</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>Data acquisition</td>
<td>0.21</td>
<td>0.16</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>0.36</td>
<td>0.30</td>
</tr>
</tbody>
</table>

*Source: Tamblyn et al., 2007.*
lum by 1920. Given the rigor of the science curriculum in the first two years of medical school and the relatively weak premedical preparation in the sciences of many applicants at that time, it is not surprising that as many as one medical student in four failed the first-year curriculum and left medical training. As discussed earlier in this chapter, this loss of students was seen as a costly “wastage,” something to be avoided if at all possible. It was for this reason that the Medical Aptitude Test (MAT) was first developed. Used in combination with grades in the premedical sciences, scores on the MAT could predict which students were at the highest risk of failing the first year of medical school.

From the beginning of the period in which the MAT was used, most medical educators fully appreciated that, while a low MAT score predicted the likelihood of failure, it was by no means 100 percent accurate. Experience had shown that substantial numbers of students who were admitted despite low grades or low MAT scores were nonetheless able to complete medical training successfully and become fully qualified physicians. When the MAT was applied increasingly to weed out low-scoring students from the admissions process, educators continued to be aware that a certain percentage of those students denied admission on this basis might have been successful in medical school. The issue was finding the most “efficient” manner in which to apply the achievement-based admissions criteria, with “efficiency” defined as attaining the optimal balance between preventing first-year failure and minimizing the rejection of otherwise qualified applicants.

Following the substantial increase in the number of applicants to medical school that came in the wake of World War II, medical educators were again concerned principally with preventing failure among admitted students. By that time, however, the failure rate of medical students had been reduced substantially. Of all U.S. medical students admitted to medical school between 1949 and 1958, only 9 percent failed to complete medical school. The first-year failure rate during this period averaged between 5 and 7 percent. Despite the markedly reduced failure rate, the use of grades in the premedical sciences and standardized test scores (by then the MCAT was being used) continued to be the principal means by which students were selected for medical school.

Beginning in the 1950s questions arose as to whether likelihood of success in the preclinical sciences taught during the first two years of medical school was either an optimal or an adequate measure. While success in the sciences was certainly important, the likelihood of success as a clinician was at least equally, if not more important. The problem was, though, that the measures used to predict success in the preclinical sciences—premedical science grades and MCAT science scores—had little if any power to predict clinical quality or skills.
Measures of Success in Medical School

Scientific Knowledge:
- NBME I and Preclinical Science Grades

Clinical Skills and Knowledge:
- NBME II, SPE, Clerkship Grades

Professional Skills:
- NBME III and Internship Evaluations

Measures of Success in Premedical Studies

Science Grades

MCAT Scores—Science

MCAT Scores—Verbal and Writing

Non-science Grades

Measures of Success in Clinical Practice

Figure 4.4. Predicting success in medical school and in clinical practice

From a series of research reports, it became apparent that clinical skills reflect a different set of attributes than scientific knowledge and that a different set of factors predict those skills. While premedical science achievement predicts success in the preclinical sciences, it is verbal ability, as measured principally by the MCAT Verbal Reasoning score, that is the strongest predictor of clinical quality. When the SPE was added to the assessment of clinical skills in the USMLE, this conclusion was reinforced. Verbal ability and other humanistic skills are the best predictors of clinical quality, especially the crucially important quality of patient communication. From these research results, it is possible to conceptualize the associations depicted in figure 4.4. The arrows in the figure reflect research about factors shown to be valid predictors, with the width of the arrow representing the strength of the prediction.

Between 2002 and 2007, the number of students graduating from medical schools nationally represented 96.7 percent of the number of students admitted to medical school four years earlier. Of the average of 3.3 percent of entering medical students who failed to graduate, approximately half experienced academic failure during the first two years of school; a substantial share of these students left medical school for personal rather than academic reasons. “Wastage” among first and second year medical students, the principal factor predicted by using premedical science grades and MCAT science scores to rank students for admission, is no longer a serious problem. Nearly every student admitted to medical school will graduate from medical school, barring personal or emotional problems. It no longer seems as
appropriate to invest achievement in the premedical sciences with the importance that admissions committees gave them beginning in 1930.

If verbal ability and humanistic skills are the principal predictors of clinical ability and professional skills, as figure 4.4 suggests, it seems only prudent to place more emphasis on measures of these abilities and skills in selecting students for medical school. As studies from the SPE suggest, these abilities and skills reflect both cognitive (i.e., academic achievement) and noncognitive (i.e., personal and psychological characteristics) aspects of medical students. In the following chapter, I examine research on evaluating the noncognitive aspects of premedical and medical students as predictors of professional success.