

A Survey of Student Assessment in U.S. Medical Schools: The Balance of Breadth Versus Fidelity

Brian E. Mavis , Bridget L. Cole & Ruth B. Hoppe

To cite this article: Brian E. Mavis , Bridget L. Cole & Ruth B. Hoppe (2001) A Survey of Student Assessment in U.S. Medical Schools: The Balance of Breadth Versus Fidelity, Teaching and Learning in Medicine, 13:2, 74-79, DOI: [10.1207/S15328015TLM1302_1](https://doi.org/10.1207/S15328015TLM1302_1)

To link to this article: http://dx.doi.org/10.1207/S15328015TLM1302_1



Published online: 20 Nov 2009.



Submit your article to this journal [↗](#)



Article views: 66



View related articles [↗](#)



Citing articles: 15 View citing articles [↗](#)

APPLIED RESEARCH

A Survey of Student Assessment in U.S. Medical Schools: The Balance of Breadth Versus Fidelity

Brian E. Mavis, Bridget L. Cole, and Ruth B. Hoppe

College of Human Medicine
Michigan State University
East Lansing, Michigan, USA

Background: Faced with the challenge to develop models of assessment relevant to work of physicians, medical schools have broadened their assessment of medical student competency.

Purpose: U.S. medical schools were surveyed to determine the extent to which student assessments have broadened beyond multiple-choice question (MCQ) examinations and preceptor ratings.

Methods: A survey mailed to 126 accredited U.S. medical schools asked respondents to indicate the frequency with which a variety of assessment methods were used in each year of the curriculum.

Results: Examinations dominated preclinical assessments. Year 3 relied heavily on faculty ratings, live observations, and MCQs. Preceptor ratings were used most in Year 4.

Conclusions: A variety of competency assessments currently are used; MCQs remain a core assessment method. Year 3 had the greatest breadth of assessment strategies. The findings suggest that educators continue to be challenged to balance the breadth of competencies sampled with the fidelity of the assessment experience.

Teaching and Learning in Medicine, 13(2), 74–79

Copyright © 2001 by Lawrence Erlbaum Associates, Inc.

Beyond their specialized knowledge, physicians must be able to apply their knowledge within the context of a clinical encounter. The appropriate use of skills and knowledge to solve clinical problems is key to defining professional competence.¹ It was the recognition that student assessment should more accurately reflect the tasks of the physician—what students can do rather than what students know—that prompted George Miller's² charge to medical educators to “just do it.”

Medical educators were challenged to develop models of assessment relevant to the work of the physician and spanning the spectrum of expected performance. Two methods historically have been used for medical student assessment: multiple-choice questions (MCQs) and preceptor ratings. MCQs provide an efficient means for testing basic abilities related to knowledge of disease and signs and symptoms, as well as basic diagnostic and treatment strategies. MCQs have clear advantages as reliable assessment tools: They can be used to test both knowledge recall and higher cogni-

tive skills, and, being easy to administer and score, they are less subject to limitations associated with content specificity.³ Conversely, the highly bounded answer frames of MCQs limit their fidelity to the actual realm of clinical work.^{4,5} Preceptor ratings, although relevant to the work of a physician, have been found to be biased and provide little discrimination of student performance.⁶ It is limitations such as these that have fueled the debate about the need for more authentic student assessments.⁷

In response, medical schools have broadened their assessment of student performance to include skills related to patient interactions, physical examinations, and problem solving.^{8,9} These assessments have taken the form of computer and live simulations, written projects, oral examinations, and other situations in which the student must create rather than choose a response¹⁰ and in which the context more closely approximates the “real world” of the physician.⁸

The current state of the art in assessing the performance of medical students has focused on the use of

Correspondence may be sent to Brian Mavis, Office of Medical Education Research and Development, A202 East Fee Hall, Michigan State University, East Lansing, MI 48824. E-mail: mavis@msu.edu

standardized patients (SPs).¹¹ A recent survey reported that 85% of U.S. medical schools used SPs for student assessment.¹² Most frequently, SP-based assessments have been implemented within the context of an objective structured clinical examination (OSCE) and used to assess students' data collection, diagnostic, and management abilities.¹³ The appeal of the OSCE is that it provides for more authentic skills assessment than paper-and-pencil testing¹⁴ as well as greater standardization than ward ratings and bedside examinations.¹⁵ The OSCE "has emerged as the best method we have of formally assessing the clinical skills of undergraduate medical students and residents" (p. 106).¹⁶

Given the many new methods that have been implemented to further competency assessment,¹⁰ it is important to ascertain how these methods are being used at present within U.S. medical schools. In this article we describe a survey of U.S. medical schools to determine which methods are in place to assess medical student performance. Of interest is the extent to which medical schools have broadened their student assessment strategies beyond MCQ examinations and preceptor ratings. Also of interest is how methodologies used to assess medical students change throughout undergraduate medical education as the curriculum moves from a focus on basic science to one on clinical medicine.

Method

A survey was mailed to 126 accredited U.S. medical schools in November 1998, to the attention of the person overseeing the undergraduate education program. Two follow-up mailings were sent to nonrespondents at approximately 30-day intervals. For the first follow-up mailing, the survey was mailed again to the attention of the person overseeing the undergraduate education program. For the second follow-up mailing, the survey was addressed to the dean of the medical school and included a cover letter requesting that the survey be forwarded to the appropriate person.

The survey, which required 15–20 min to complete, was developed around a list of specific assessment methods:

- MCQs.
- Oral examinations.
- Essay questions, including written exams, papers, and essay projects.
- Laboratory practical examinations.
- Review of written records and chart notes.
- Faculty or preceptor ratings of students individually or in small groups.
- OSCEs.
- Other standardized patient assessments not included as part of an OSCE.
- Live observation of students.

- Logbooks.
- Rated or graded case presentations.
- CBX (a specific pilot program) or computer-based testing (excluding MCQs).
- U.S. Medical Licensing Examination step examinations.
- National Board of Medical Examiners (NBME) subject examinations.
- Other.

Respondents were asked to indicate the frequency with which each method was used for student assessment during the undergraduate medical curriculum. Each year of the curriculum was presented in a grid format, and respondents were asked to rate each student assessment method using the following rating scale: A = 1–2 times; B = 3–12 times; C = 13 or more times; D = not used.

Results

Eighty-nine medical schools (70%) returned a survey, of which 87 (68%) were completed and used for this summary. No significant differences were found when responding and nonresponding medical schools were compared on the basis of whether the medical school was public, private, or community based. We also compared responding and nonresponding schools using data from the 1998 Institutional Goals Ranking Report.¹⁷ No group differences were identified on the basis of federal research grant and contract dollars received, the percentage of graduates planning to practice in state, the percentage of underrepresented minority graduates, and the proportion of graduates who assumed faculty positions. When we compared the proportion of graduates who chose primary care specialties, we noted that the difference between responding (32.0%) and nonresponding (29.3%) schools approached statistical significance, $t(121) = 1.93, p = .057$.

Figure 1 illustrates the proportion of medical schools reporting the use of each assessment method at least once at some point during the undergraduate medical curriculum. More than 80% of responding schools reported using almost all of the assessment methods included in the survey, creating an impression of uniformity across schools in the use of multiple methods to assess medical students. Traditional assessments such as MCQs, preceptor ratings, laboratory practical examinations, written record reviews, live observation, and case presentations remain the dominant modes of assessment. The only method used by fewer than half (46%) of the schools was computer-based assessment.

Table 1 presents a summary of the ratings for each assessment method for each year of the curriculum and highlights the variability of medical student assess-

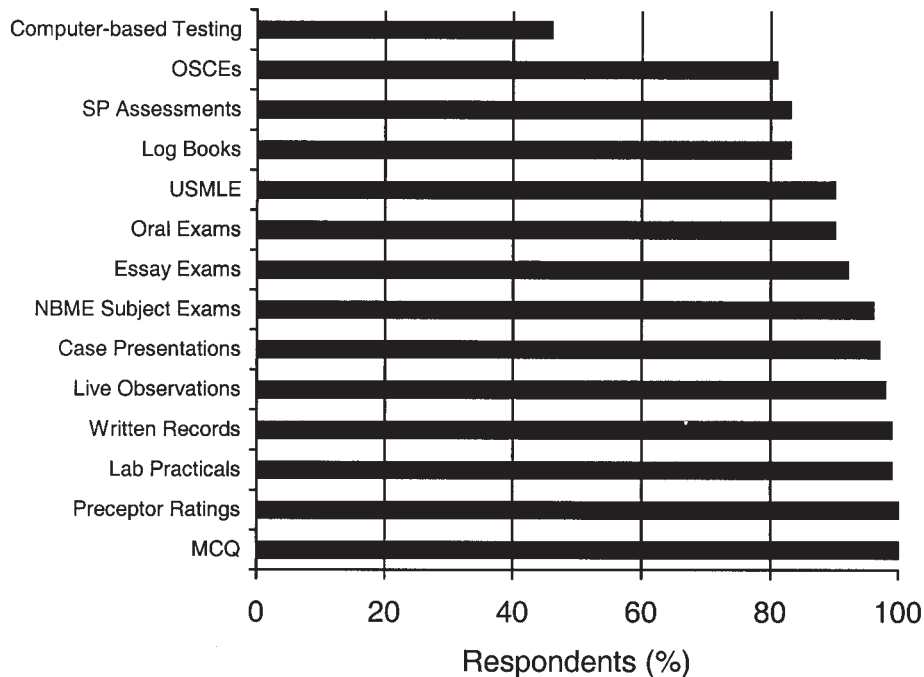


Figure 1. Proportion of medical schools reporting any use of each student assessment method. OSCE = objective structured clinical examination; SP = standardized patients; USMLE = U.S. Medical Licensing Examination; NBME = National Board of Medical Examiners; MCQ = multiple-choice question.

ment across schools. The boldface type in Table 1 indicates the ratings of 50% or more of the respondents for a specific assessment modality, within a specific year. The pattern of boldface type in Table 1 suggests what is typical of medical student assessment. The different methods of student assessment have different rates of use across the 4 years of the curriculum. The data reported in Table 1 illustrate specific changes in use of each assessment strategy over time.

Preclinical Curriculum

Examinations dominate assessment in the preclinical curriculum. School-based MCQ examinations were used by almost two thirds of the schools more than 12 times in each of the first 2 years, and NBME subject examinations were used by one third of responding schools 3–12 times during each year. Laboratory practical examinations were used frequently during the 1st (83%) and 2nd (63%) years. Half of the schools also reported the use of essay questions for medical student assessment in Years 1 and 2.

Faculty or preceptor ratings and live observations of students were reported throughout the curriculum, including at least three times per year in Years 1 and 2 for most schools. Methods based on live observation that are more standardized, such as OSCEs and SP encounters, were used less frequently in Years 1 and 2. OSCEs were reportedly used at least once in Year 1 by 25% of schools and by 50% in Year 2, although 75% and 50%

of schools reported no use during these same years, respectively. Almost half (45%) of the schools reported some use of SPs for assessments outside of an OSCE in Year 1; this increased to 63% in Year 2.

Clinical Curriculum

In the clinical curriculum, Years 3 and 4 were less similar to each other than Years 1 and 2 were to each other. Year 3 was characterized by heavy use of faculty ratings, live observations, and MCQ examinations. MCQ examinations in Year 3 were among the dominant modes of assessment, although less frequently than in Years 1 and 2. NBME subject examinations were used 3–13 times in Year 3 by 75% of the respondents. Case presentations, written records, and logbook assessments all had their heaviest use in Year 3, presumably on required clinical clerkships. Again, standardized forms of assessment such as OSCEs were used less frequently, and nearly 40% of schools reported no use at all in Year 3.

In Year 4, the use of MCQ-based assessment decreased dramatically. One third of schools indicated that they did not use them at all in Year 4, and only 23% used them more than twice. Most schools (88%) reported no use of NBME subject examinations in Year 4. Preceptor ratings were the dominant mode of assessment, with nearly all schools reporting some use, typically multiple times during the year. Forty-one percent of the schools reported using preceptor ratings

Table 1. Summary of Ratings Related to the Frequency of Medical Student Assessment by Curricular Year (Percentages)

Assessment Modalities	Year 1				Year 2				Year 3				Year 4			
	Not Used	1-2 Times	3-12 Times	> 12 Times	Not Used	1-2 Times	3-12 Times	> 12 Times	Not Used	1-2 Times	3-12 Times	> 12 Times	Not Used	1-2 Times	3-12 Times	> 12 Times
MCQ	0	2	34	64	0	2	36	62	0	11	83	6	33	44	22	1
Oral Exams	72	22	6	0	63	26	11	0	18	44	37	1	58	26	12	4
Essay Questions	17	26	51	6	23	26	50	1	48	33	19	0	61	26	13	0
Lab Practical Exams	1	11	83	5	19	17	62	2	94	4	3	0	98	1	1	0
Written Records	66	20	13	1	35	24	35	6	4	5	49	43	15	6	41	38
Preceptor Ratings	1	25	52	9	5	17	68	10	2	0	42	56	0	2	57	41
OSCE	75	20	5	0	50	39	10	1	39	41	19	1	73	24	3	0
SP Assessment	55	27	17	1	37	40	21	2	52	34	13	1	79	20	1	0
Live Observations	31	30	33	6	17	19	54	10	4	7	47	42	11	11	41	37
Logbooks	78	18	4	0	70	20	10	0	28	20	48	4	58	18	23	1
Case Presentations	54	25	20	1	37	23	35	5	7	16	50	27	24	13	49	14
Computer-Based Testing	78	12	10	0	73	17	10	0	81	14	5	0	91	5	4	0
USMLE	99	0	0	1	20	78	2	0	59	30	11	0	39	60	1	0
NBME Subject Examinations	61	6	33	0	36	28	36	0	25	0	75	0	88	6	6	0

Note: Boldface type indicates 50% or more of respondents for a given assessment method within 1 year of the curriculum. MCQ = multiple-choice question; OSCE = objective structured clinical examination; SP = standardized patient; USMLE = U.S. Medical Licensing Examination; NBME = National Board of Medical Examiners.

more than 12 times during Year 4. Again, live observations, case presentations, and written record evaluations were the other methods frequently used in Year 4.

Conclusions

Recent curricular innovations in medical education have fostered an interest in increasing the breadth of methods used for student assessments. Our data bear this out. MCQ-based examinations, in the form of class examinations; NBME subject examinations; and licensure examinations remained the most frequently used approaches for assessing student performance during the first 3 years of the curriculum. These examinations allow for coverage of many topics and clinical scenarios in a short time, which is important for drawing conclusions about student competence.³ Such examinations also are less costly when compared with many other methods. The use of NBME subject examinations appears to be a trend among many medical schools. According to the 1999 Association of American Medical Colleges Graduation Questionnaire, 89% of medical school graduates nationally reported having taken an NBME subject examination, an increase from 77% among 1995 graduates.¹⁸ Whether this is in response to less time available by clinical faculty to write MCQ examinations, or whether the quality of these examinations is more defensible against student complaints, is not known. The goodness of fit between the content of these subject examinations and local curricular emphasis remains a concern, along with their relatively low fidelity to clinical situations.

At most medical schools, faculty and preceptor ratings were the most widely used source of student performance assessment across the 4 years of the curriculum. Given that these methods have been found to be susceptible to rater bias and offer little discrimination among students,⁶ it is positive that Year 3 is the point in the curriculum at which there is the greatest breadth of student assessment methods. The survey results suggested that although there was still a reliance on assessment methods with low fidelity (MCQ testing) or low sensitivity (preceptor ratings), many schools reported the use of other, more standardized methods for assessing knowledge and skills.

Most schools reported using OSCEs for student assessment, most often during the 2nd or 3rd years of the curriculum. In addition, many medical schools also reported other assessments based on the use of SPs, implemented throughout the first 3 years of medical school. This is consistent with findings from the Association of American Medical Colleges¹² report that 85% of medical schools used various SP-based assessments. The authors of the report concluded that although SP-based experiences have become commonplace for medical student instruction and assessment, they are less frequently used for decisions

related to student advancement. Others¹⁹ have confirmed this finding.

Some assessment methods remain closely linked to a specific curricular context. Laboratory practical examinations are an important form of assessment during what typically is considered the preclinical curriculum, in which the focus is on basic science education; laboratory practical examinations were seldom used during the last 2 years of medical education. To a lesser extent, essay question examinations showed a similar pattern: They were most likely to be used during the preclinical curriculum. Conversely, written record review and case presentations were used most often during the clinical (3rd and 4th) years of the curriculum. Oral examinations were most likely to be used as part of 3rd-year student assessments.

Virtually all medical schools reported the use of live observations of students, with the number of schools reporting the use of this method increasing from Year 1 to Year 3 of the curriculum. This is consistent with findings from the Association of American Medical Schools, which reported similar data based on a survey of medical school administrators by the Liaison Committee for Medical Education. However, similar surveys of medical school graduates, in which 27% of respondents said they did not have their clinical skills evaluated by faculty observation, cast doubt on these findings.²⁰ This discrepancy might be the result of differences between students and administrators in how terms such as *faculty observation* were defined. Just as likely is that it could represent the difference between the written policy statements and actual faculty practice regarding medical student assessment.

Limitations

There are a number of limitations regarding the results reported in this study. The first is related to the response rate. Although a majority of medical schools are represented among the respondents, almost one third of the medical schools did not return a completed survey after numerous attempts to solicit participation. There were no significant differences on most of the indicators comparing respondents and nonrespondents. The one difference that was identified, regarding the proportion of graduates in primary care specialties, although statistically significant, does not seem to be of pragmatic significance. Nonetheless, as with any survey, there is a possibility of other unmeasured sources of bias differentiating respondents and nonrespondents.

Another limitation of the survey might reflect the difference between policy and practice. The survey was directed to administrators responsible for the undergraduate medical curriculum. In completing the survey, respondents provided an overview of the student assessment practices at their school. The extent to which his or her responses accurately reflected the ac-

tual assessment practices of faculty are unknown. One study already cited²⁰ found that data from multiple sources might not always lead to the same conclusion. When multiple sources are not available, it is difficult to determine the validity of the findings.

Summary

The survey illustrates the increasing breadth of assessment methods used to determine the competence of medical students. Reliability and validity continue to be major concerns in the implementation of assessments; however, the broader challenge of defining competency has focused attention on the need to balance the competing demands of fidelity of assessment and breadth of competencies sampled for assessment.²¹ In light of this challenge, a strategy for competency assessment that incorporates a variety of assessment methods remains the optimal approach.²² The survey results indicate that although a broad array of assessment methods are being implemented throughout the curriculum, more traditional methods, such as MCQs and preceptor ratings, remain central to student assessment.

At best, the results of this survey represent the collective intentions of medical school faculty with regard to assuring the competency of medical students; however, these results are limited to what was reported by administrators. What was reported on paper might not reflect how student assessments are implemented in practice. Neither does the implementation of a student assessment method imply that the resulting information will be used for summative purposes, such as promotion or graduation. Nonetheless, the survey serves to document the continuing efforts to address the challenge of developing assessment that is more consistent with the tasks of a physician, with an eye toward "what a student can do."

References

1. Kane M. The assessment of professional competence. *Evaluation and the Health Professions* 1992;15:163–82.
2. Miller G. Conference summary. *Academic Medicine* 1993;68:471–4.
3. Fowell S, Bligh J. Recent developments in assessing medical students. *Postgraduate Medical Journal* 1998;74:18–24.
4. Ferland J, Dorval J, Levasseur L. Measuring higher cognitive levels by multiple choice questions: A myth? *Medical Education* 1987;21:109–13.
5. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: An alternative to multiple-choice tests in assessing physicians' competence. *Academic Medicine* 1999;74:539–46.
6. Albanese M. Rating educational quality: Factors in the erosion of professional standards. *Academic Medicine* 1999;74:652–8.
7. Linn R, Baker E, Dunbar S. Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* 1991;20(8):15–21.
8. Newble D. Assessing clinical competence at the undergraduate level. *Medical Education* 1992;26:504–11.
9. Swanson D, Norman G, Linn R. Performance-based assessment: Lessons from the health professions. *Educational Researcher* 1995;24(1):5–11.
10. Van der Vleuten C. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Science Education* 1996;1:41–67.
11. Van der Vleuten C, Swanson D. Assessment of clinical skills with standardized patients: The state of the art. *Teaching and Learning in Medicine* 1990;2:58–76.
12. Association of American Medical Colleges. Emerging trends in the use of standardized patients. *Contemporary Issues in Medical Education* 1998;1(7):1–2.
13. Tamblyn R, Barrows H. Data collection and interpersonal skills: The standardized patient encounter. In A Tekian, C McGuire, W McGaghie (Eds.), *Innovative simulations for assessing professional competence* (pp. 77–104). Chicago: University of Illinois, 1999.
14. Vu N, Barrows H. Use of standardized patients in clinical assessments: Lessons from the health professions. *Educational Researcher* 1994;23:23–30.
15. Colliver J, Swartz M. Assessing clinical performance with standardized patients. *Journal of the American Medical Association* 1997;278:790–1.
16. Reznick R, Blackmore D, Dauphinee D, Smee S, Rothman A. The objective structured clinical examination (OSCE) of technical skills: The Canadian experience in licensure. In A Tekian, C McGuire, W McGaghie (Eds.), *Innovative simulations for assessing professional competence* (pp. 105–12). Chicago: University of Illinois, 1999.
17. Association of American Medical Colleges. *Institutional goals ranking report*. Washington, DC: Author, June 24, 1998.
18. Association of American Medical Colleges. *Graduation questionnaire all schools reports*. Washington, DC: Author, 1999. Retrieved July 17, 2000 from the World Wide Web: <http://www.aamc.org/meded/gq/archives/>
19. Mavis B, Cole B, Hoppe R. *New ways of assessing students: Old ways of determining grades*. Paper presented at the annual meeting of Research in Medical Education, Association of American Medical Colleges, Washington, DC, October 1999.
20. Association of American Medical Colleges. The role of faculty observation in assessing students' clinical skills. *Contemporary Issues in Medical Education* 1997;1(1):1–2.
21. Norcini J. Measurement issues in the use of simulation for testing professionals: Test development, test scoring and standard setting. In A Tekian, C McGuire, W McGaghie (Eds.), *Innovative simulations for assessing professional competence* (pp. 195–211). Chicago: University of Illinois, 1999.
22. Newble D, Dawson B, Dauphinee D, et al. Guidelines for assessing clinical competence. *Teaching and Learning in Medicine* 1994;6:213–20.

Received 7 August 2000

Final revision received 11 October 2000