



VitalSigns

The Challenge of Assessing Clinical Performance

Office of Medical Education Research and Development

Spring 1997

The Challenge

This edition of *VitalSigns* focuses on clinical performance assessment, an issue at the center of discussions nationally and of special significance to the mission of our medical school.

Clinical performance assessment provides evidence that graduates can function in clinical settings. While debate surrounds *why* and *how*, there is little consensus on *what* it is. In practice, performance assessment moves beyond tests of what students know to records of what students can do.

The debate on *why* has focused on concerns that assessments historically have concentrated on knowledge, and the recognition that knowledge by itself is not sufficient for competent medical practice. This has been formalized by the inclusion of performance assessment in the LCME accreditation standards, as well as plans for skill-based assessments by the NBME.

Any discussion of *how* to conduct student assessment necessarily raises questions of how to sample skills and abilities; the reliability, validity, and costs of assessment; and how the data will be used to make decisions. These questions are no less pertinent in the discussion of how to assess clinical performance.

Are CHM Students Clinically Competent?

Despite major resources devoted to clinical performance, the College of Human Medicine cannot give assured answers to such questions as “How good is the clinical performance of CHM students at graduation?” and “Does the clinical curriculum of the college assure competent performance by CHM graduates?” While extensive, the clinical performance assessments of the college lack an agreed standard of performance that defines the meaning of assessment results.

Because of the exceptional attention CHM gives to clinical communication and interaction, it was early dubbed “the college of the bedside manner.” The current curriculum includes 216 preclinical hours of teaching interviewing, physical examination, and patient education. CHM’s unique Integrative Clinical Correlations links basic science instruction to clinical reasoning. The problem-based instruction in year two is increasingly evaluated with clinical simulations. CHM’s 60 weeks of community-based required clerkships (four weeks above the U.S. mean) include the distinctive “Clinical Medicine in the Community,” “Core Competency,” and linked primary care experiences. Clinical performance is monitored throughout, from clinical science courses to the clerkships’ performance ratings (*How CHM Assessment Stacks Up* on page 3 of this issue). CHM has even secured ratings of graduates’ performance by residency directors (PGY-1) and parallel self-ratings by graduates in the second post-graduate year.

“As one watches a man handle a patient it is easy to tell whether or not he has had a proper training, and for this purpose fifteen minutes at the bedside are worth three hours at the desk.”
William Osler, M.D. 1849-1919

Performance rating forms dominate the assessment of clinical performance of CHM students and graduates. These ratings share in a chronic “inflation” problem (*That Pesky Lake Wobegon Effect*, p. 6). Clinical performance ratings of CHM students and graduates are always “acceptable” or better. Without a process for defining expected behaviors and for standardizing ratings, there is no assurance that ratings that “exceed expectations” signify performance that exceeds (or even meets) the basic standards of the CHM faculty.

CHM faculty have not defined what is ultimately expected in clinical performance from the curriculum as a whole or from each curriculum block. Most CHM clerkships have not fully stated their performance expectations or assured a reliable check on performance vis-à-vis expectations. Consequently, assessment outcomes such as average clinical performance ratings or clerkship honors often vary erratically from year to year and community to community.

Continued on page 8

Grads Rate CHM Clinical Skills Curriculum

Time Spent on Clinical Skills Rated Favorably

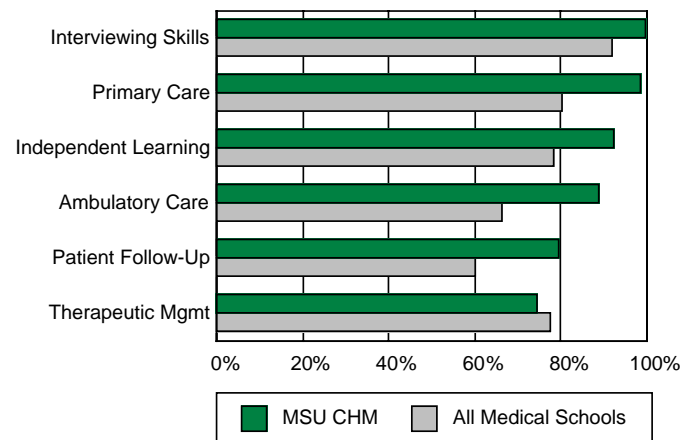
New CHM graduates rated their instruction time as “appropriate” more often than graduates nationally in most areas related to clinical performance, including care of the elderly, diagnostic skills, clinical decision-making, teamwork, communication skills, and the doctor/patient relationship. In the two areas where CHM graduates reported time spent as “inadequate”—therapeutic management and care of hospitalized patients—the difference between CHM and national numbers was less than 5%.

Consistent with our primary care mission, most graduates felt enough time had been spent on both patient interviewing skills and primary care. Also, reports of time spent on ambulatory care and patient follow-up were higher at CHM than nationally. CHM’s emphasis on professional behavior and lifelong learning is reflected in ratings of time spent on independent learning and self-evaluation.

Although CHM compares favorably to other schools, with differences of up to 22 percentage points on some dimensions, there is still room for improvement. One-fifth of MSU graduates indicated that more instructional time should be devoted to patient follow-up, and 25% felt that time spent on therapeutic management was inadequate.

Overall, CHM graduates perceived fewer problems in the curriculum than their counterparts nationally. Follow-up surveys of CHM graduates validate these findings: the majority reported that they were at least as well prepared as their peers in areas related to clinical performance.

Graduating Students Who Rate Time Spent as at Least “Appropriate”



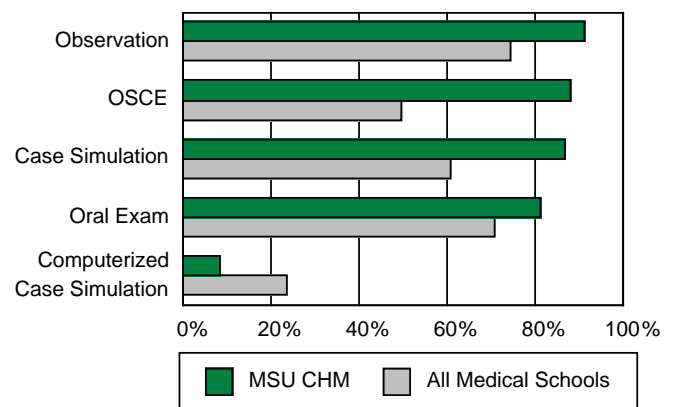
Source: Association of American Medical Colleges, 1987-1996, and Michigan State University College of Human Medicine, 1990-1996

Students Report on Clinical Skills Assessment

Graduating CHM students were more likely to report that their clinical skills had been evaluated in the context of patient care, either on simulated or actual patients, than were graduating students nationally. CHM is more likely than other schools to use performance assessment methods, such as OSCEs using standardized patients, case simulations, faculty observation of an H&P, and standardized patients for interviewing or physical exams.

These results indicate that many of the required elements are in place for a high-fidelity evaluation of students’ clinical skills; however, they need to be organized into a cohesive picture of what students can actually do. The success of this system is predicated on assessments that are linked directly to specific curricular objectives and that are psychometrically sound.

Strategies Used to Assess Clinical Skills, as Reported by Students



Source: Association of American Medical Colleges, 1995-1996

PERSPECTIVE: Ruth B. Hoppe, M.D., Associate Dean

When one examines activities at CHM being directed at assessment of what students DO (as opposed to what they KNOW), one is impressed with the faculty effort. This effort is remarkable when one considers the complexity and geographic dispersion of our clinical campus. We have lots to be proud of, including a seminal role in the field of clinical simulation and assessment through the efforts of such CHM pioneers as Ray Helfer, M.D., Jack Maatsch and their colleagues.

But, we are no longer leaders in the field. The use of standardized patients is now routine in the nation's medical schools, and pre-graduation OSCEs have been adopted by over a third. The accreditation standards of the LCME now require performance assessment for all schools. Beyond this, there is our responsibility as educators. Broadened assessment will be required as the public demands more accountability. As learning environments become more diverse, more remote and more challenged by costs, we will need to assess that our educational goals are being met in all settings.

To reexamine and refocus our system of performance assessment, we must first accept our *roles and responsibilities*: every medical educator must equip him/herself to design and conduct performance assessment that meets peer-reviewed standards of quality. Also, we must address the issue of *balance*. Given the three "legs of the medical education stool," we over-assess one (knowledge), under-assess another (skills) and generally ignore the third (professionalism). Together we must formulate a more balanced assessment system. This system must move from the descriptive to the evaluative, use clear

How CHM Assessment Stacks Up



observed H&P



PBA



simulated pt. interview



oral exams



oral case presentation



interview clinical write-ups

ratings

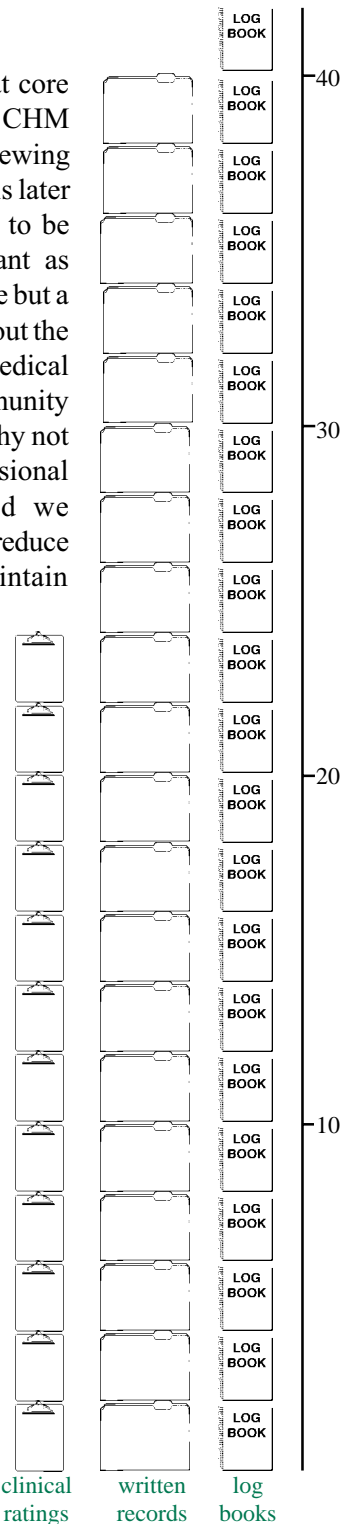
written records

log books

standards, and give the most "bang" for our limited assessment "buck."

We must address key questions. What core competencies must be required of all CHM graduates? We assess beginning interviewing skills early - what happens to these skills later when there are real clinical problems to be solved in limited time? As important as physical examination skills are, they are but a part of the physician's toolkit. What about the ability to critically access and assess medical evidence? To consider issues of community health as well as individual illness? Why not begin to routinely assess the professional behavior of *all* students? Should we collaborate with our sister institutions to reduce effort and cost? How can we maintain standards if we diminish the number of knowledge assessments? Can we clinicians fulfill our faculty responsibilities in assessment at a time when we are all pressed to be more clinically productive?

In my view, the immediate future of medical education is all about assessment. It is my hope that we can build on our long-standing commitment in this field, reassert our role as leaders nationally, and build a student assessment system that is envied from afar. There are many challenges ahead, but also some wonderful opportunities. We have the talent. Let's DO it!



CHM students are observed and assessed frequently during their first three years. The log book is a popular tool for monitoring the 42 procedures students are expected to perform during required clerkships.

Dimensions of PBA

Questions suggesting the dimensions by which the quality of a performance-based assessment (PBA) might be considered are listed below. While not exhaustive, they sample issues which affect the quality of the experience presented in the PBA.

SELECTION

Relevance: To what extent are the behaviors to be evaluated ones which would be used in a clinical setting?

Appropriateness: Are the abilities assessed and performance expectations appropriate given the level of training of the student?

Integration: To what degree does the assessment focus on the integration of basic abilities rather than on the basic skills themselves?

Fidelity: In what ways does the context of the assessment event simulate a real clinical encounter?

METHODOLOGY

Standardization: How uniform is the assessment experience across students, examiners and clerkship communities?

Cuing: How do the problem presentation and response options compare to those that occur in clinical settings?

Reliability: Are the judgments made about student performance consistent across examiners or raters?

IMPLEMENTATION

Remediation: For assessments of skill mastery, are opportunities for remediation available?

Ease: Are the resources required to implement the assessment commensurate with the types of judgments to be made?

Evaluation: How do the assessment data contribute to a system of ongoing curricular feedback and development?

What is Performance Assessment? Why is CHM Doing It?

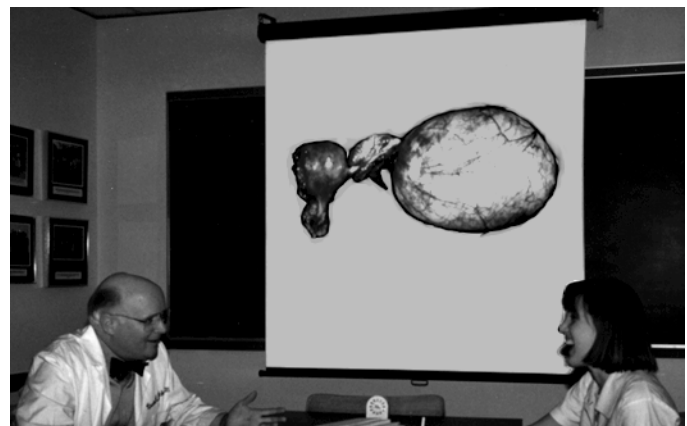
These questions were the focus of a half-day retreat held in December. The meeting was a rare opportunity for faculty from all curricular blocks to discuss how we currently gather evidence of student performance and how the quality of this evidence might be improved.

Ruth Hoppe opened the meeting with thoughts on why CHM should invest in assessments focused on student performance. **Rebecca Henry** elaborated a framework for discussion, presenting ten dimensions by which the authenticity of performance-based assessment (PBA) could be appraised. These dimensions addressed the selection of assessment events, methodological considerations and implementation issues, and are presented on the left. **Karen Ogle** used the dimensions to discuss the Block II OSCE, formerly a part of the Clinical Skills Program. Afterwards, **Jane Turner** led a panel discussion of how the dimensions apply to current Block II and Block III assessments: **Kathy Lovell** explained the Block II Neuromuscular PBA; **Maddy Dodson** described the Ob/Gyn Oral Examination; and **Ashir Kumar** reported on the Pediatric Clerkship experience.

From the retreat emerged a number of collective concerns as well as issues unique to particular curricular components. Of common interest were concerns about the integration of PBA within the existing curriculum, resources for developing and implementing PBAs, reliability and standardization, educating faculty about PBA and methodologies for assessing complex integrative skills. Preclinical faculty questioned the utility of assessing performance and integration when students were still learning fundamental skills and concepts. Clinical faculty pointed to the tremendous variability in quality and use of data sources for decision-making about students' clinical performance.

Participants found the meeting provided faculty with an opportunity to learn how each Block addresses performance assessment and to consider strategies for improving existing assessments. The meeting also generated interest in further PBA development and the possibility of a system of clinical performance assessment across the CHM curriculum.

Bruce Drukker,
Ob/Gyn Chair,
administers an
oral exam (see
p. 5)



Neuro PBA Yields Feedback for Students and Curriculum

CHM faculty created a performance-based assessment (PBA) to evaluate students' abilities to integrate the neurologic content of the problem-based learning (PBL) curriculum with clinical skills. The assessment was organized around cases in which localization of a neural lesion required use of both basic science knowledge and clinical skills. The objective was to provide second-year medical students with helpful, low cost feedback about their performance.

The neurologic PBA included three stations linked by a case: a computer-based case description along with six questions designed to stimulate hypothesis development; performance of a focused neurologic examination on a simulated patient; and a write-up of an assessment and plan. Two cases were created, with half of the of the class acting as "doctor" and half acting as "patient" for each case. Assessment results provided students with feedback about their performance, but were not included in clinical skills or PBL course grades.

Students reactions to the challenges and benefits of the integrative PBA experience were positive. Faculty

received valuable information on the ability of students to recall concepts and skills and to apply them in a specific clinical setting. The information indicated the effectiveness of parts of the curriculum and helped to generate ideas for enhancing student learning in the future. Results also stimulated faculty discussion about remediation experiences for students who did not perform at a satisfactory level.



A Block II student formulates hypotheses for a computer-based neurologic case.

Ob/Gyn Oral Exam Targets Diagnosis and Management

Twenty-five years ago the Department of Obstetrics and Gynecology initiated an oral examination. Faculty believed that the written exam, though useful, did not provide sufficient information about how students would diagnose and manage real clinical problems. The oral exam has three components: recognition and interpretation of fetal monitor tracings, ultrasound and pathologic or anatomic material; an obstetrical case; and a gynecologic case.

Each examination is conducted by two faculty members who independently score the student on data acquisition, problem solving, management, and knowledge. "Our department emphasizes standardization. A strict set of procedures guide the selection of the questions, how the questions are asked and scored," notes Maddy Dodson, who directs medical education for the department.

Conducting the oral exam is as stressful for the department as it is for students. The exams require a great deal of faculty resources. On average the department will spend

about 50 faculty hours in face-to-face evaluation and another 25 hours in report writing for each rotation. Training new examiners is difficult and includes a variety of techniques including videotaped exams and direct observation of the oral. The effort in training for standardization is rewarded, as reliability is very good.

"Just about every year we weigh the pros and cons of the oral examination and decide that it is just too good as a measure of student performance to toss out," Dodson concludes. One student recently wrote on his clerkship evaluation that he would never have believed how worthwhile the oral exam proved to be and suggested that it be expanded. Students report that the mock oral offered in each community helps them prepare for the 45-minute exam that requires them to think on their feet. The department has considered using computers instead of faculty but the faculty continue to respond that interacting with the students reveals important information about their skill levels.

Assuring Professional Competence: What's Required?

In the United States, the medical profession won a hard-fought trust from legislators and the public in medical schools' ability and willingness to assess the competence of its potential graduates for professional practice. The professionalization of medicine in the U.S. would not have been possible without explicit, rigorous control over the standards for medical education. These standards include the development, use, and review of credible assessment systems.

Legitimate assessment should produce evidence that graduates can function as competent physicians. Assessment *systems* are required, for no single method of assessment captures the diverse forms in which competence should be evident. One means for deciding whether an assessment system is "enough" is to consider how well it avoids projecting a distorted or oversimplified picture of professional competence. For example, assessment focusing exclusively on memory of medical knowledge misrepresents medical practice. The criterion of "authenticity" promotes selection and construction of assessment tasks that, as a whole, vividly portray relevant central, representative tasks of clinical work. If this criterion is met, the assessment system can promote learning, by informing students and teachers about what clinical performance demands.

"A healthy evaluation system is one way to honor the trust that society places in physicians."

Daniel Hunt, M.D. 1992

A basic demand of trustworthy assessment is capturing judgments that are consistent. Meeting this criterion of *reliability* means that we can trust that a judgment of a student as "superior" on a clinical task would be replicated by other faculty and in other settings.

Assessment should measure what it is intended to measure. This criterion of *validity* is met if assessment: appears, on the face of it, to measure what it is supposed to measure, as when students and faculty accept an oral examination requiring recognition of fetal monitor tracings as a reasonable representation of an OB student task; predicts future behavior (e.g., performance of second-year medical students predicts future success in conducting focused neurological examinations); and elicits empirical data that fit theoretical models, as when data elicited

with Quality of Life scales correspond to our understanding of issues related to life quality.

Standards used to assess competence should inspire confidence. The development, use, and scrutiny of assessment methods constitute part of an expected dialogue with the public, the profession, and the individual practitioner-graduate. In this dialogue, assessment can make explicit what professional competence means.

That Pesky Lake Wobegon Effect

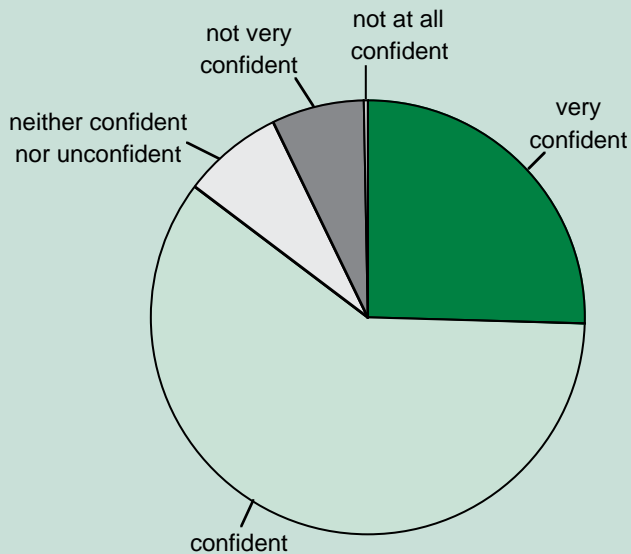
Average scores on CHM's Clinical Performance Evaluation (CPE) forms are unrealistically high. Although a standardized CPE seems well-suited for systematic feedback about student performance, in practice these scores are often inconsistent with other measures and observations. As CPE is the single means available for judging the clinical performance of students across clerkships, it's vulnerable to distortion. Despite efforts to modify the CPE form and to orient faculty to the problems resulting from inflated scores, the college's average CPE score consistently falls in the "superior" range, averaging over 7.5 on a 9-point scale. It is difficult to accept that superior performance is really the norm among our students.

Why does score inflation occur? Inflated averages on performance ratings are not rare in medical education or in other professions education—CHM is not alone here. This phenomenon is most likely the product of many complex factors: how the faculty perceive their rating task, the system of medical education which has become litigious and sometimes adversarial, and pressure from students to have high scores for the residency match.

However, a more serious implication of score inflation is that the evaluation is not truthful to students. When students who are experiencing difficulty are identified for assistance, all too often they point to CPE scores as counter

Continued on page 7

New Graduates Confident in their Clinical Skills



Source: Association of American Medical Colleges, 1992-1996

Continued from page 6

evidence. Clerkship directors are on thin ice when faced with such contradictory evaluations. Even more problematic is when preceptors raise concerns orally about a student but then don't substantiate them with a comparable rating. "Recently the faculty reviewed the progress of a student on our clerkship. They agreed the performance was not satisfactory but the CPE scores could not support the assessment," offered Kim Anderson, Director of Undergraduate Education for the Department of Surgery.

Steps have been taken to improve CPE ratings. The surgery clerkship instituted several changes to address the score inflation problem: they no longer release to faculty scores on the final exam until the CPE is submitted; instruction to faculty describes the problem; and faculty are "given permission" to use average ratings on students—something that was seldom done.

Despite its flaws, the clerkship CPE ratings still provide unique information. These ratings predict residency directors' evaluations of graduates' performance in the first year of residency in a way that is not captured by any other assessment method. Also, the CPE is the only performance measure that is common to all clerkships, producing longitudinal data about student progress.

Psychosocial Skills Related to Sustained Career Satisfaction

Trust in the medical profession comes in part from objective assessment. But physicians' self-assessments also offer a vantage for seeing if physicians' themselves derive satisfaction in practices for which their medical school training has prepared them.

At CHM, alumni rate their satisfaction with their career in a survey conducted ten years after their graduation. Seven CHM cohorts, representing graduates of the classes from 1980 through 1986, have responded to the ten-year survey. Eleven elements defined career satisfaction including satisfaction derived from providing service, conducting research, interacting with colleagues, being independent, and having a controllable lifestyle. Factors considered in relation to career satisfaction included features of CHM undergraduate training, residency training, speciality, debt load, patient characteristics and practice profile.

The medical and psychosocial skills preparation provided by the medical school, as well as primary care specialty, medical school community training site, and time spent teaching emerged as significant predictors of career satisfaction ten years after graduation. Preparation provided by CHM continued to affect the satisfaction of alumni with their established medical careers. These aspects of training associated with career satisfaction reflect dimensions emphasized in the College of Human Medicine's mission to provide competent, humanistic, and primary-care oriented physician training.

Oops!

The wrong e-mail address for comments was listed in the first issue of *VitalSigns*.

Suggestions, reactions, challenges, and questions about this issue or the first issue are welcome.

By e-mail:

OR

Send letters to:

*vitalsig@
pilot.msu.edu*

VitalSigns
OMERAD
A202 East Fee Hall
Michigan State University
East Lansing, MI 48824

Continued from page 1

Meaning can be anchored by tying assessment to a widely accepted standard. Steps in the U.S. Medical Licensure Examination provide that standard for assessments of knowledge. However, NBME has not yet made similar standards available for clinical performance, although it has been exploring the use of standardized patients in national licensure. Performance-based assessments, because of their inherent face validity, can also be used to anchor the assessment system, but only if they are frequent and varied enough to assure reliability and comprehensiveness.

The number of performance-based assessments applied in the College is increasing. The challenge to the CHM faculty is twofold: to define a coherent network of expectations and measures for clinical performance and to design a set of strategically located performance assessments to anchor the network. Without this, there will be no way to answer the question, “Is performance of CHM graduates acceptable?”

Grapevine

Thanks to all of you who sent in comments on the first issue of *VitalSigns*, some of which appear below. Keep those cards and letters coming.

“The first issue of *VitalSigns* was provocative and made me think about our tradition of primary care. Seeing other schools increase their production of primary care grads makes me wonder if we might lose our identity. While there are many factors influencing career choice, I think about our circumstances: do we have a culture that supports primary care? Do we have the curriculum to meet our mission? What tools are needed by future physicians? If we step back and address these questions, it might be interesting to compare the ideal curriculum and environment with what we currently have.” *Henry Barry, M.D., Chair, CHM Curriculum Committee*

“Thanks for *VitalSigns!* I am delighted to have a publication that so succinctly demonstrates our progress toward realizing our mission. It seems that we so often fail to document what we do well. This publication will assist me in educating others about who we are and what we do. The review of the specialty choice data from the AAMC matriculation survey is particularly meaningful to those of us active in admissions: the anonymous post-matriculation survey shows that being upfront with the mission of the college during the admission process seems not to be counter-productive as feared by some.” *Christine Shafer, M.D., Co-Chair of the CHM Admissions Committee, CHM Class of ’79*

“Your last *VitalSigns* was an excellent report on the state of CHM primary care training outcomes. I would be interested in hearing more about the last two years of clinical training and factors in the undergrad/GME interface which help/hinder the choice of a generalist training program. Also, don’t forget the growing number of Med/Peds programs: over 75% of their residents go into primary care practice. It’s important to preserve a variety of training options to maximize choices and entice students into generalist practice.” *Laura Carravallah, M.D., Program Director, Combined Med/Peds Program, Hurley Medical Center, CHM Class of ’89*

*Office of Medical Education Research & Development
Michigan State University, A217 East Fee Hall
East Lansing, MI 48824-1316*

This Issue of *VitalSigns*:

Clinical Education Unvalidated	1
Grads Give CHM High Ratings	2
PERSPECTIVE of Assoc. Dean Hoppe	3
What is Performance Assessment?	4
Neuro PBA in Block II	5
Ob/Gyn Oral Exam	5
Warranting Professional Competencies	6
The Lake Wobegon Effect	6
Alumni Reflect on Training Impact	7