



Accreditation Council for Graduate Medical Education

GUEST EDITOR'S INTRODUCTION

The Accreditation Council for Graduate Medical Education publishes the **ACGME Bulletin** four times a year. The Bulletin is distributed free of charge to more than 12,000 individuals involved in residency education, and is also available on the ACGME's Web site (www.acgme.org) for viewing and printing. The ACGME receives and publishes letters to the editor in the interest of furthering dialogue about accreditation, program quality and matters of general interest in residency education. Inquiries, comments or letters should be addressed to the editor.

Pamela Derstine, PhD
Guest Editor

Ingrid Philibert
Editor

515 North State Street
Suite 2000
Chicago, Illinois 60610-4322

Phone 312.755.5003
Fax 312.755.7498
Web: www.acgme.org

Improvement Through the Application of the Competencies

Turning the Challenge into Opportunity

Since 2001, program directors have been challenged to define specific objectives, provide learning opportunities, and develop evaluation systems that support and assess resident learning in the six competency domains that – by now – we all know by heart: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, systems-based practice. Over the past four years, programs have been asked to use resident performance data for program improvement. Beginning July, 2006, the accreditation focus will be on evidence that programs are making data-driven improvements, using not only resident performance data, but also external measures.

“Beginning July, 2006, the accreditation focus will be on evidence that programs are making data-driven improvements, using not only resident performance data, but also external measures.”

With this special issue of the Bulletin, we present a variety of examples, both large and small and from different specialties, to provide program directors, GME committees, DIOs and faculty with a broad perspective for identifying improvement opportunities. We present two essays that provide a framework for consideration. David Leach, MD, ACGME Executive Director, urges us to base improvement on the “EUJA cycle” by asking us to consider four questions and letting the answers guide our efforts. Arianne Teherani, Assistant Professor of Medicine at UCSF, reminds us that professionalism is at the heart of both individual resident and program performance and improvement efforts should be based in this competency, addressing the admissions process, resident evaluation, program evaluation, and institutional commitment.

Reports from four specialties – anesthesiology, psychiatry, emergency medicine, and internal medicine – provide examples of the use of data to improve the selection of residents. David Metro and Joseph Talarico describe a fairly straight-forward use of available data on interview scores and resident performance to guide not only their admissions process, but also their evaluation process during residency education. Karon Dawkins and her colleagues retrospectively compared pre-residency applicant evaluations to

Improvement Through the Application of the Competencies

Guest Editor's Introduction:

Turning the Challenge into Opportunity

P. Derstine, PhD1

Executive Director's Column:

Using Assessment for Improvement: It Begins with Experience

D.C. Leach, MD4

How Can We Improve the Assessment of Professionalism Behaviors in Graduate Medical Education?

A. Teherani, PhD5

Improving the Resident Selection Process

Your Admissions Process Can Help You

D.G. Metro, MD, J.F. Talarico, DO7

The Relationship between Psychiatry Residency Applicant Evaluations and Subsequent Residency Performance

K. Dawkins, MD, R. D. Ekstrom, MA, MPH, A. Maltbie, MD, R.N. Golden, MD9

Designing Instruments to Assess Residency Applicants

G. Bandiera, MD, Med, FRCPC1111

A Postgraduate Orientation Assessment at the University of Michigan

J.C. Janus MD, S.J. Hamstra, PhD, L. Colletti, MD, M. L. Lybson, MD13

Improving the Curriculum

Practice-Pattern-Based Curriculum Development

R.S. Isaacson, MD15

Journal Club: A Tool to Teach, Assess, and Improve Resident Competence in Practice-based Learning and Improvement

A.G. Lee, MD20

Morbidity and Mortality Conference: A Practice-Based Learning Tool for the Performance Improvement of Residents and Residency Programs

J.C. Rosenfeld MD, MEd, FACS23

Improving the Teaching of Competence

Toward a Broader Understanding of Interpersonal and Communication Skills in Resident Education

B. Joyce, PhD26

The Effect of Standardized Patient Feedback in Teaching Surgical Residents Informed Consent

K. Reed, DO, FACOS29

Asking to Learn: A Study of Performance Outcomes for Program Improvement

G. R. Bergus, MD, MA and M. Emerson, MA, JD30

Improving Assessment

The Development and Application of a Competency-based Assessment Tool

L. M. Reich, MD33

Fostering Self-Assessment and Self-Directed Learning in the Intensive Care Unit

A.S. Clay, MD36

Assessing the ACGME Competencies with Methods That Improve the Quality of Evidence and Adequacy of Sampling

R.G. Williams, PhD, G. L. Dunnington, MD38

Improving Remediation

Planning for Remediation

S.A. Schartel, DO43

Reluctance to Fail Poorly Performing Residents – Explanations and Potential Solutions

N. Dudek, MD, MEd, M. Marks, MD, MEd & G. Regehr, PhD45

A Multi-Faceted Approach to Resident Evaluation and Remediation

P. M. Boiselle, MD, B. Siewert, MD48

Remediation in Graduate Medical Education: A Point of View from Surgery

C.G. Rehm MD, P.A. Rowland PhD51

RRC/IRC Column and Other News from the ACGME

.....53

RRC Development Course Amis to Enhance Reparation for Review of the General Competencies

.....54

Improving Patient Outcomes – the Link to Resident Performance

The July Phenomenon: Fact or Fiction

S. H. Borenstein, MD, J.C. Langer, MD54

A Study of Resident Supervision in the Operating Room

K.M.F. Itani, MD, FACS, S.F. Khuri, MD, FACS56

Teaching about Transparency: Linking Resident Performance and Patient Care Outcomes

T. Foster, MD, MPH, MS, K. E. George, MD, M. R. Lauria MD, MS58

post-residency evaluations for already graduated residents. They use five dimensions (empathic quality, academic potential, clinical potential, team player, and overall rating) in order to determine how effective the admissions process was in predicting resident performance and identifying opportunities for improvement. Glen Bandiera describes two instruments that he and Glenn Regehr developed for assessing applications and scoring interviews. In the process, he provides an excellent tutorial for identifying the characteristics of good evaluation instruments and suggests reasons for the difficulty in predicting resident performance. Finally, Jennifer Janus and her colleagues describe the “POA”, a postgraduate orientation assessment intended to establish the baseline abilities of first-year residents at the start of residency.

Richard Isaacson describes a structured approach to curriculum improvement that is based on both resident and hospital needs and sound principles of education. He then applies the model to internal medicine, surgery, and neurology/radiology. More modest curriculum improvement efforts are described by Andrew Lee (ophthalmology) and Joel Rosenfeld (surgery). Both are efforts to improve resident performance in outcomes related to practice-based learning and improvement: a transformed Journal Club for teaching and assessing principles of evidence-based medicine; a restructured Morbidity & Mortality conference that not only improved resident performance but also identified curriculum weaknesses in other competency areas that resulted in program improvements.

Barbara Joyce helps us to reconceptualize the competency of interpersonal and communication skills as it applies to resident education, providing a description of the seminal literature and summarizing areas of focus appropriate for the medical, surgical and hospital-based specialties. Kendall Reed then describes the program he and his colleagues developed to improve surgical residents’ abilities to obtain informed consent. Communication skills important for lifelong learning were addressed by George Bergus and Myra Emerson (family medicine) by testing the questions, “What characteristics of questions seeking clinical consultations promote physician learning?” and “Do resident skills for asking effective clinical consult questions improve as they progress in their training?”

We introduce the area of improving the assessment of residents with a description of a competency-based assessment tool, CEPI (comprehensive educational performance improvement), developed by Lawrence Reich. It takes into account the professional needs and desired outcomes of residents, the faculty, the program and the institution. The author provides specific examples of its use in a didactic small-group workshop, outpatient clinical experience, and evaluation of resident performance by ward nurses. Alison Clay focuses on improving resident self-assessment and self-directed learning in critical care medicine by developing multiple assessments and a mentored resident portfolio. Assessment by direct observation can be improved by attention to the quality of evidence and the number of ratings. Reed Williams and Gary Dunnington describe how they improved their use of this

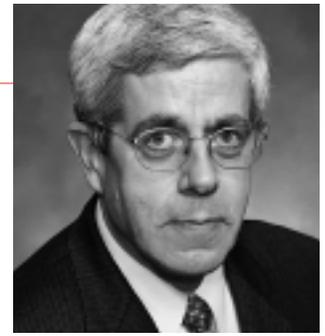
important method: develop the operative performance rating system, use the patient assessment and management examination process, and improve the semi-annual program resident evaluation meetings by the use of resident portfolios.

An important and often overlooked area for program improvement is remediation, already introduced by Janus et al. with their use of POA at the beginning of residency education. Scott Schartel provides an in-depth overview of important aspects to consider when examining your own remediation programs. Nancy Dudek and her colleagues summarize their important findings on why faculty “fail to fail” poorly performing residents – in particular, the lack of remediation opportunities – and the implications for the resident, faculty, programs and institutions for meeting this challenge. A system for linking resident evaluation and remediation in a radiology program is described by Phillip Boiselle and Bettina Siewert, stressing the concept of creating an alliance between the resident, program director and faculty educational liaison with specialized experience in the resident’s area of difficulty. Two examples demonstrate the effectiveness of the model. The impact of remediation programs offered through the American Board of Surgery for residents unsuccessful at board certification is described by Christina Rehm and Pamela Rowland, concluding with a recommendation to incorporate remediation within residency programs.

Can programs link resident performance outcomes to patient care outcomes? We offer three examples to stimulate your creative energy. Steven Borenstein and Jacob Langer asked if pediatric surgery patients experienced suboptimal care during the month of July as a consequence of the influx of new first-year residents. They found that care was not affected and identified ways the program and faculty compensated to achieve this favorable outcome. Kamal Itani and Shukri Khuri utilized an extensive national database of surgical outcomes to determine if the ACGME requirement for increasing levels of resident responsibility throughout residency has an impact on patient outcomes, finding that outcomes were not affected and highlighting adjustments made by the program. The decision by her program to participate in contributing to an obstetrical database, recording information about all deliveries, facilitated a program improvement initiative by Tina Foster and colleagues. The initiative not only educated residents in the use of quality improvement metrics, but also sought to directly relate the outcomes of inpatient care with their own and their team’s practice. These are early but promising examples addressing the connection between patient outcomes and resident performance.

With this collection of examples “from the field” we hope you see the challenge of making data-driven program improvements to meet ACGME accreditation requirements as an opportunity to examine the fine work you are already doing and make it even better. ■

Pamela Derstine, PhD, is Senior Project Manager in the ACGME’s Department of Research and Education.



Using Assessment for Improvement: It Begins with Experience

David C. Leach, MD

Improvement gurus advocate something called a cycle of: “plan; do; check; and act” or PDCA (sometimes called a PDSA cycle wherein “check” is replaced by “study”). They offer the PDCA cycle as a tool and a discipline for continuous quality improvement. While sympathetic with improvement gurus, I must admit that I do not like this model. It invites us to begin in our heads with a mental model of planning. Rather, I favor a much more ancient approach that asks us to begin with experience. After all, we have a sensory apparatus, why not use it?

“A simple disease model becomes complex as patients present with multiple interacting acute and chronic diseases and as science demonstrates that yesterday’s dogma is no longer true.”

The formation of residents is embedded in the experiences of patients – poignant, compelling, sometimes dramatic and sometimes mundane experiences. Resident formation depends on the substrate of the human condition: birth; disease; suffering; death – the routine work of medicine. While much of medicine is supported by good evidence and widespread agreement about what should be done, most of medicine is not. It is laden with uncertainty and lacks solid evidence and agreement about how to proceed. A simple disease model becomes complex as patients present with multiple interacting acute and chronic diseases and as science demonstrates that yesterday’s dogma is no longer true. In such a world how does one improve? As programs assess the resident’s competence in the six competencies, what data helps? What correlates with good learning? How can residents and those helping them use assessment data for improvement? How can we help residents become better doctors? As messy as it is, improvement begins where the resident lives: in the experiences encountered in the daily care of patients.

Four questions may simplify and organize attempts to assess for improvement: what can be done to improve the resident’s experience; what can be done to improve the resident’s understanding of his or her experience; what can be done to improve the judgments residents make based on their experiences and their understanding of their experiences; and what can be done to improve the actions residents take having judged and decided what to do?

The resident’s experience can be improved by improving patient care. Residents’ direct participation in patient care makes it impossible to separate the quality of their experience from the quality of patient care. If patient care is shabby, their formation will be shabby; if it is excellent, they can learn excellence. Improving patient care improves resident formation. Assessing and telling the truth about the quality of patient care helps. While patient care is at the heart of the work, it is sometimes hard to determine the quality of patient care at a given institution. Hospital websites frequently are filled with statements about “excellent care,” “top ten hospitals,” “best doctors and best technology,” yet few websites have real data. Fortunately there are exceptions. On the public section of the Dartmouth Hitchcock website are data about institutional performance with regard to several diseases and procedures. Dartmouth data are side by side with national data (top 10% and national average). Telling the truth is a first step. Accurate data about institutional performance enables the resident’s experience to be grounded in something other than the illusions created by public relations departments.

Understanding may be enhanced by reflection. As residents try to make sense out of their experiences they use conversations with peers, with faculty, with nurses, with

“Dartmouth data are side by side with national data (top 10% and national average). Telling the truth is a first step. Accurate data about institutional performance enables the resident’s experience to be grounded in something other than the illusions created by public relations departments.”

anybody to clarify what has actually happened. Journaling helps; recalling specific events of the day can consolidate the lessons of the day. Most residents lack time for formal reflection. However, its importance makes it a top priority, eventually resulting in reflection-in-action with its attendant time- and quality-saving features.

As residents attempt to assess and understand their experiences and patient care outcomes, autopsies can help. Even in this age of sophisticated technology there is a discordance rate of 40% between the clinical diagnoses of

the cause of death and the postmortem diagnoses findings at autopsy.¹ Yet the frequency of autopsy has declined and many residents lack this important feedback on their thinking.

Critical incidents offer another and compelling opportunity for reflection. The internal conversations that clarify what happened; the frequently difficult conversations with patients and their families; and the identification of system issues that contributed to the incident are all deeply memorable and formative.

Assessing the art of medicine may be more helpful than assessing knowledge of the science of medicine. While knowledge of the science supporting medicine is essential it is not sufficient. It is in the nature of science to ever more closely approximate the truth without actually getting there. Scientific knowledge is constantly changing. In contrast, the skills associated with the “art” of medicine – getting a good history, performing a careful exam, listening deeply, thoughtfulness in action, reflection on experience – are enduring and foster closer participation in the direct experience of patient care.

Judgment may be improved by both art and science. Science applies to universals; art is always unique. Science informs us about several aspects of pneumonia: its natural history; bacteriology; antibiotic sensitivities; etc. Art helps us understand what to do when Mary Smith gets pneumonia. Good judgment is informed by her uniqueness as well as the offerings of science. Paying attention to the art of medicine improves resident formation. Assessment tools that help give feedback about the art of medicine include videotapes of interviews with patients or their surrogates, direct observation of the resident’s encounters with patients by experienced clinicians, patient feedback about the resident’s performance, resident’s self-assessment, chart review, etc.

Residents get things done. They must learn not only the body of knowledge in their specialty but also the frequently arcane necessary steps (the “locally useful knowledge”²) needed to get things done. **In taking action residents encounter a new curriculum;** there are good and less good ways of taking action. Execution of a decision can go smoothly or not.

When it comes to resident formation I would replace the PDCA cycle with EUJA (experience, understanding, judgment and action).³ Assessment of actions, judgment, understanding and experience offers a balanced portfolio of data for resident formation to improve. When coupled with the six competencies the experience-oriented model offers a learning plan for life. ■

¹ Low tech autopsies in the era of high tech medicine. Lundberg, G. JAMA, 280, 14, 1273-4, 1998.

² Paul Batalden, personal communication. Dr. Batalden has developed the concept of “locally useful knowledge” as part of his model of the “science of clinical practice” and as a companion science to the science of disease biology.

³ Residents are not the only ones who learn from conversation. The EUJA model is a modified version of one described by Jack Boberg, PhD, former Executive Director of the Surgery RRC in a conversation in September, 2005. It reflects the work of Bernard Lonergan and, to some extent, David Kolb, Stephen Brookfield and Donald Schon.

How Can We Improve the Assessment of Professionalism Behaviors in Graduate Medical Education?

Arianne Teherani, PhD

Respect. Compassion. Integrity. Commitment to excellence. Sensitivity to patients. Responsiveness to patient diversity. These are a few of the professionalism competencies the ACGME Outcome Project mandates residents should demonstrate. The ACGME will soon require programs to demonstrate that resident assessment data is used to improve resident performance and programmatic educational effectiveness. Yet, what are the best ways to measure these competencies? Perhaps more importantly, how does one stratify the importance of these competencies? Lastly, are there data to support this new ACGME requirement?

Although we do not have the answers to these questions, we will make an argument for the importance of measuring professionalism behaviors based on our research. In 2004, we showed that unprofessional behavior in medical school was

“Three domains of unprofessional behavior emerged that were significantly related to later disciplinary outcome. These domains were: 1) poor reliability and responsibility, 2) lack of self-improvement and adaptability, and 3) poor initiative and motivation.”

associated with disciplinary action by a state licensing board decades later.¹ We examined medical student files of 68 physicians who had graduated from University of California, San Francisco (UCSF) and had been disciplined by the Medical Board of California and compared them with 196 matched UCSF graduates who had not received state disciplinary action. We learned that those disciplined were three times more likely to display unprofessional behavior in medical school. Upon discovering this, we then asked which types of behaviors were the most predictive of future disciplinary action.

In a pilot study we re-examined the physicians’ student files from our first study to isolate behaviors associated with later disciplinary outcomes.² Types of unprofessional behavior were qualitatively analyzed using a research-based, comprehensive in-house instrument and subsequently quantified to determine if any association existed between each behavior and disciplinary action. Three domains of unprofessional behavior emerged that were significantly related

to later disciplinary outcome. These domains were: 1) poor reliability and responsibility, 2) lack of self-improvement and adaptability, and 3) poor initiative and motivation. We then conducted a similar but larger study of three medical schools and 40 state licensing boards.³ Once again, students who displayed unprofessional behavior in medical school were three times more likely to be disciplined years later by a licensing board. We were also able to confirm two of the three domains of unprofessional behavior. Those who were most irresponsible in medical school were nearly nine times more likely to receive subsequent disciplinary action, and those who had severe problems with lack of self improvement and adaptability were three times more likely to have future disciplinary actions.

We propose that effective assessment of professionalism at the graduate medical education level depends on a multidimensional program infrastructure. Programs need to be able to identify individuals in need of intervention, develop and implement individual learning plans, and measure outcomes. A robust infrastructure that would support these needs must include all of the following: 1) admission documentation, 2) learner assessment, 3) program evaluation, and 4) institutional commitment.

“Great strides have been made to standardize the secret language of the Medical Student Performance Evaluation (MSPE) or dean’s letter, yet variation exists in meaning from institution to institution. Are there other mechanisms by which residency programs can collect more information about applicants?”

1. Admission Documentation

The admission process for residency education has focused predominantly on the academic standing of those applying. Great strides have been made to standardize the secret language of the Medical Student Performance Evaluation (MSPE) or dean’s letter, yet variation exists in meaning from institution to institution. Are there other mechanisms by which residency programs can collect more information about applicants? A variety of documentation, exercises, and assessments could be developed and tested to be included in the application process to better understand the applicant. Examples of such material might include validated instruments which measure critical aspects of one’s reliability and ability to self improve. Exercises focused on assessing interpersonal skills and professional behaviors such as clinical cases and standardized patient exams, or standardized checklists or rating scales of professionalism behaviors to be completed by those familiar with the applicant are just a few useful assessment tools.

2. Learner Evaluation

Assessment in the clinical setting has earned a reputation for ambiguity and inconsistency among both learners and educators. A periodic review of the system in place and a needs assessment by faculty would reveal if the system provides accurate and appropriate data about unprofessional behaviors or if the implementation of new assessment tools is necessary. The inclusion of measurement tools within existing assessment modalities is often the most efficient way to ensure that professionalism can be measured. Items can be included about unprofessional behaviors on clinical evaluation forms (e.g., documentation of how often a resident was absent or tardy) or routine clinical assessments (e.g., items documenting residents’ response to faculty feedback during standardized patient exams). Critical to the success of an evaluation system is the existence of a centralized structure that facilitates the consolidation and review of multiple evaluations for each resident. An essential part of this centralization is the preparation of faculty to look for appropriate and inappropriate behaviors, especially those of particular concern (such as identification of the resident heavily resistant to feedback). In parallel, developing and implementing a longitudinal curriculum around professionalism and ethics will educate residents about the importance of recognizing these problematic behaviors in themselves and others. The latter provides a much needed complementary approach to faculty development. Ultimately, all such data must be collected in a manner that is efficient so that at-risk individuals can be identified with adequate time for feedback and, if necessary, intervention.

3. Program Evaluation

A systematic and routine evaluation of the professionalism evaluation system in place is essential to its success. A key component of this evaluation system is periodic review of the learner evaluation system. Faculty opinions on the effectiveness of the instruments in use as well as the system in place are critical, especially when addressing the sensitive and important issue of professionalism. Residents’ assessment of the utility and consistency of the evaluation system and tools will provide rich information about its fairness and accuracy. An essential element of program evaluation, particularly complex when assessing professionalism, is the tracking of graduates in practice. Periodic surveys and/or interviews with graduates, the collection or use of data collected by other organizations (such as the American Medical Association), and communication with state licensing boards can yield information that is vital to understanding how well a program was able to detect those individuals at risk.

4. Institutional Commitment

The value an institution places on professionalism will permeate through the institutional milieu, affecting residents, faculty and, fundamentally, the evaluation process itself.⁴ Any benefit to residents accrued from the new ACGME initiatives will be undermined unless the parent institution articulates its

commitment to professionalism and acts on that commitment. The residency education program cannot ask the residents to be more professional than the faculty. Faculty has to be held accountable for their professionalism, and that translates into a respectful educational environment with appropriate emphasis on education. Just as reducing the work week to 80 hours does not guarantee improved educational satisfaction,⁵ neither will a mandate for professionalism unaccompanied by a scrutiny of faculty professionalism. As faculty members are trained to recognize unprofessional behavior in residents, residents should also be educated to identify unprofessional behavior in faculty. Residents should be given ample opportunity to report faculty behavior of concern within a safe and confidential environment. Most of all, institutions should develop an approach to respond to and rectify ongoing professionalism issues.

Professionalism is a complex and controversial competency that is often difficult to assess. A recent surge of research in this area has contributed to understanding the importance of unprofessional behavior during medical school and the types of “red-flag” behaviors linked to later disciplinary action. The mandate that programs use assessment data to improve learner and program outcomes coincides with the timeliness of the evidence-based outcomes of professionalism research. This opportune timing can pave the way for a better and more accountable assessment system for professionalism – cultivating an understanding of professional behavior not only for faculty but also residents. ■

Arianne Teherani, PhD, is Assistant Professor in the Department of Medicine and Office of Medical Education, and Assistant Director for Evaluations in the Office of Educational Technology at the University of California, San Francisco.

The author gratefully acknowledges the guidance provided by Maxine Papadakis, MD, and Mary Banach, PhD.

- ¹ Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad. Med.* 2004; 79: 244-9.
- ² Teherani A, Hodgson CS, Banach M, Papadakis MA. Domains of unprofessional behavior during medical school associated with future disciplinary action by a state medical board. *Acad. Med.* 2005; 80(10 Suppl): S17-20.
- ³ Papadakis MA, Teherani A, Banach MA, Knetter TR, Rattner SL, Stern DT, Veloski JJ, Hodgson CS. Disciplinary action by medical boards and prior behavior in medical school. *N. Engl. J. Med.* 2005; 353(25): 2673-82.
- ⁴ Suchman AL, Williamson PR, Litzelman DK, Frankel RM, Mossbarger DL, Inui TS. Toward an informal curriculum that teaches professionalism: Transforming the social environment of a medical school. *J. Gen. Intern. Med.* 2004; 19: 501-504.
- ⁵ Vidyarthi AR, Katz PP, Wall SD, Wachter RM, Auerbach AD. Impact of reduced duty hours on residents' educational satisfaction at the University of California, San Francisco. *Acad. Med.* 2006; 81(1): 76-81.

Your Admissions Process Can Help You

David G. Metro, MD, Joseph F. Talarico, DO

We are all looking to make our residency programs the best possible. Traditionally we look mainly at the curriculum for improvements. What is often overlooked is the resident selection process itself. An integral characteristic of a successful program is the selection of residents that fit the program. In the Department of Anesthesiology at the University of Pittsburgh, we decided to evaluate our selection process to determine if it could help us improve the program.

A traditional selection process

Our process begins by determining which applicants will be selected to interview after applications are received through the Electronic Residency Application Service (ERAS). Information reviewed includes USMLE scores, medical school performance, personal statement, dean's letter, and letters of recommendation as well as any personal communication we have received (**Table 1**). The formal interview process consists of dinner with residents, four to five interviews with faculty, lunch with more residents and a tour of the facilities. Upon completion of the interview, each faculty interviewer ranks the applicants on a 1–5 Likert scale, using the same criteria that are used in resident evaluation

Table 1
Areas Evaluated with Admission Interview

- Personality aspects
- Enthusiasm, energy, liveliness
- Drive, motivation to excel
- Assertiveness
- Warmth, friendliness, humor
- Poise, composure
- Maturity
- Intellectual aspects
- Intelligence and problem solving skills
- Articulatness, communication skills
- Critical or analytical skills
- Quality of answers
- Quality of questions
- Strength and appropriateness of goals
- Knowledge of anesthesia as a specialty
- Knowledge of Pitt program
- Comment on strengths and weaknesses
- Compare with present average resident
- Overall evaluation

At the completion of the interview process, our entire Resident Education Committee (approximately 50 members) meets to evaluate each applicant using their ERAS information, scores given by faculty interviewers, recommendations and resident input. Using these criteria, each member and resident representative rates each applicant with a score from 1 (undesirable) to 5 (outstanding) based on their suitability for admission. This is done using a closed ballot system. Scores from each rater are treated identically regardless of their role in the residency program. Once all applicants are rated, the average score for each applicant is calculated and listing the candidates in order of their average score forms the rank list.

“Using these criteria, each member and resident representative rates each applicant with a score from 1 (undesirable) to 5 (outstanding) based on their suitability for admission. This is done using a closed ballot system.”

Evaluating the process

We analyzed two years of our selection data to see if it could predict performance in residency. Looking at current residents in the program who had gone through the application process, we evaluated two areas of resident performance: in-service training exam (ITE) scores and daily evaluation scores. (Table 2) We thought that ITE scores would not correlate since many things other than knowledge base were looked at in our selection process. However, since the criteria for the interview process and resident daily evaluation were similar, we felt these scores would have strong correlations.

Table 2

Daily and Year-end Evaluation Criteria

- Knowledge (basic science and clinical, facts, applications)
- Judgment (data collection, evaluation, problem solving, decision making)
- Motor skills (regional anesthesia, general anesthesia, monitoring)
- Pre-anesthetic assessment and planning (patient evaluation, anesthetic plan, presentation to faculty)
- Work habits (organized, neat efficient, punctual, charting)
- Intrapersonal attitudes (courteous, cooperative, reliable, accepts criticism, communicates well, good rapport)
- Intrapersonal attributes (motivated, industrious, flexible, ethical, takes initiative, functions well under stress)
- Overall impression

We found that there was no significant correlation between selection committee scores and subsequent resident evaluation scores related to any of the evaluation criteria. There was a statistically significant correlation between selection committee scores and PGY-2 ITE results.

We could explain these results in a number of ways. First, the selection process itself eliminated most, if not all, of the unqualified candidates prior to entering our residency. Perhaps our process was successful in eliminating people who would not succeed, but could not select those who would excel. It also suggests that our faculty continue to take USMLE scores into higher regard than other factors when ranking an applicant. From an evaluation standpoint, there is a chance that our evaluation process itself cannot distinguish fine differences between residents. Factors within the residency program also could contribute to the lack of correlation, as residents who are having difficulty are more likely to take advantage of special opportunities, such as tutoring, to improve their evaluations.

A new selection process

Since this review of data, we have begun looking for ways to improve. The interview process itself is no longer seen as a time to evaluate medical knowledge and intellectual suitability, but as a time to see who will fit best in the program from a personality standpoint. We use the interview day to give applicants as much information as possible about the program, and have them meet as many faculty members and residents as possible so that they can make the best selection for their future learning. We limit the number of interviewers so that an applicant's standing within the interview year is more apparent. We have also changed our scale for this process to 1-9 rather than 1-5 in order to further distinguish differences among candidates, and are now using the general competencies as the basis for the interview evaluation scores. Faculty receives training on how to interview and evaluate applicants before each interview season. This itself becomes an annual discussion on qualities we are looking for in residents as well as how to improve the process.

We have also begun reexamining our evaluation process during residency education as a result of this review. The evaluation of residents is one of the most important aspects of a residency program and consistency of evaluation data gathered from multiple faculty at multiple sites is an area of concern. Consequently, we have implemented an online evaluation form that specifically addresses each of the competencies. Each faculty member who works with our residents fills out these evaluation forms online. When a resident fills out their case log on our departmental web site, this triggers a reminder email to the faculty member listed for that day. The evaluation form itself is in the form of the

Table 3

Daily Evaluation Based on the Competencies

1-2-3 Unsatisfactory	4-5-6 Satisfactory	7-8-9 Superior
<ul style="list-style-type: none"> • ABA Essential attributes • Patient Care – “Residents must be able to provide patient care that is compassionate, appropriate, and effective for the treatment of health problems and the promotion of health.” • Medical Knowledge – “Residents must be able to provide patient care that is compassionate, appropriate, and effective for the treatment of health problems and the promotion of health.” • Professionalism – “Residents must demonstrate a commitment to carry out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population.” • Communication and Interpersonal skills – “Residents must demonstrate effective information exchange and teaming with patients, their families, and other health care professionals.” • Practice-Based Learning and Improvement – “Residents must be able to investigate and evaluate their patient care practices, appraise and assimilate scientific evidence, and improve their patient care practices.” • Systems Based Practice – “Residents must demonstrate an awareness of and responsiveness to the larger context and system of health care, and the ability to effectively call on system resources to provide care that is of optimal value.” 		

general competencies (Table 3). The competencies are defined on the form and a 1–9 Likert scale is used to achieve greater discrimination of resident abilities. Faculty discuss the attributes and rating scale during faculty meetings with the goal of achieving a common understanding resulting in more reliable evaluation data. We are also exploring other areas of program improvement, including consistent evaluators, performance in simulation, 360-degree evaluations and use of QI data.

In order to be successful, one must be willing to change as needed. Only by continually evaluating the process and its success in its stated goal can one really know if the objective is being consistently accomplished.

David, G. Metro, MD, is the Associate Program Director, Department of Anesthesiology, University of Pittsburgh School of Medicine.

Joseph F. Talarico, DO, is the Chair, Residency Evaluation and Competence Committee, Department of Anesthesiology, University of Pittsburgh School of Medicine.

The Relationship Between Psychiatry Residency Applicant Evaluations and Subsequent Residency Performance

Karon Dawkins, MD, R. David Ekstrom, MA, MPH, Allan Maltbie, MD, Robert N. Golden, MD

The psychiatry program faculty at the University of North Carolina (UNC) invests significant time and resources interviewing and selecting residents because we believe that a resident without the requisite skills or commitment can be more damaging to a program than an unfilled slot. Even when programs can identify desired qualities, the ability to predict performance can be faulty. Most residency program directors can recall stellar applicants who performed poorly, and others who presented less well on paper and in person who turned out to be outstanding. Program directors seeking to improve their programs may find our study of the predictive value of our selection process a useful example.

We examined the residents who matched for academic years 1995-96 through 1998-99.¹ Their pre-residency applicant evaluations (10 point rating scale for each of five dimensions (empathic quality, academic potential, clinical potential, team player, and an overall rating) were compared to their post-residency evaluations. The matriculated residents for each year were rank ordered and then divided into thirds, which were labeled as “A”, “B”, and “C”. This “match third” served as our primary, pre-residency measure of expected performance. At the end of each four-year residency cycle, they were evaluated again. The Psychiatry Education Office prepared the files of these graduates by temporarily removing all pre-application

“There was no significant association between pre-residency selection evaluations and post-residency final evaluations in any of the four cohorts. When the four cohorts were combined, there was no significant association (Fisher’s exact test: p = 0.38, k = 0.19, agreement = 43%).”

materials so that final ratings could be done independent of the original evaluation materials. The top third were assigned an ‘A’, the middle third a ‘B’, and the remaining third a ‘C’. After independent evaluations were completed, the three raters met, discussed any evaluation discrepancies, and arrived at a consensus evaluation for each resident.

There was no significant association between pre-residency selection evaluations and post-residency final evaluations in any of the four cohorts. When the four cohorts were combined, there was no significant association (Fisher’s

exact test: $p = 0.38$, $k = 0.19$, agreement = 43%). Using Cochran-Mantel-Haenszel analysis to control for year, the hypothesis of non-zero correlation between pre- and post-residency evaluations could not be rejected ($p = 0.077$). The Spearman correlation between these measures was 0.24 ($p = 0.07$). Weighted kappa was $k = 0.20$, and the 95% confidence interval for the kappa statistic (-0.004 to 0.40) included zero, indicating that the association between evaluations did not differ significantly from that which would be expected to occur by chance.

“These findings suggest that our selection process was no better than chance at identifying the future top performers, contrary to our initial hypothesis. Sensitivity was about 42% (8/19) for those applicants originally deemed outstanding and highly desirable, 42% (8/19) for the middle tier, and 45% (9/20) for the third tier applicants.”

These findings suggest that our selection process was no better than chance at identifying the future top performers, contrary to our initial hypothesis. Sensitivity was about 42% (8/19) for those applicants originally deemed outstanding and highly desirable, 42% (8/19) for the middle tier, and 45% (9/20) for the third tier applicants. The overall results of this study of four cohorts of residents indicated only a small, non-significant association between pre- and post-residency evaluations.

Potential confounds include the absence of post-residency evaluation data on interviewed applicants who did not match at UNC. This pool included applicants we ranked highly as well as those low on our rank list. Another potential confound was our decision to compare post-residency cohorts with each other and force a uniform distribution. It could be that some groups were uniformly outstanding, or uniformly average. An additional limitation was the relatively small sample size. The raters' past knowledge of and working relationships with these graduating residents could lend itself to bias, or greater accuracy, in the post-residency evaluations.

Whether our findings could be generalized to other programs is not known, however similar studies might be done. There were no problem residents in the sample, so the selection process may have been successful to that degree. Objective data has improved since this study. Dean's letters are less able to suppress variables like failing or marginal grades, leaves of absence, or repeating an academic year. However, program directors cannot assume that other, more subjective qualities they would seek to avoid have already been screened out or remediated in medical students.

While we utilize objective measures (e.g., minimum USMLE performance) in deciding whether to offer an interview, the interview itself remains an important means to assess interpersonal and communication skills. To decrease inter-rater variability, the preliminary rank order list is now based on the evaluations of those faculty members who have had interviews or other interactions with all of the applicants, as opposed to faculty who have only evaluated a subset of the applicants. We use the evaluations of a core group of seasoned faculty interviewers, as well as a small group of resident interviewers, to modify the preliminary rank order list. There are now opportunities for applicants to interact with

“There are now opportunities for applicants to interact with the residents in an informal setting. While we do not solicit a formal evaluation from this gathering, the residents know that they can offer feedback as appropriate.”

the residents in an informal setting. While we do not solicit a formal evaluation from this gathering, the residents know that they can offer feedback as appropriate. We hope that the combination of objective and subjective assessments, utilizing better defined criteria and tools, will improve our ability to predict future performance. ■

Karon Dawkins, MD, is the Psychiatry Program Director and Associate Professor, R. David Ekstrom, MA, MPH, is a retired Research Assistant Professor, Allan Maltbie, MD, is Director of the Psychiatric Consultation Service and Professor, and Robert N. Golden, MD, is Vice Dean, School of Medicine and Professor; all in the Department of Psychiatry at University of North Carolina, Chapel Hill.

¹ Dawkins K, Ekstrom RD, Maltbie A, Golden RN. The relationship between psychiatry residency applicant applications and subsequent residency performance. *Academic Psychiatry*. 2005; 29(1): 69-75.

Designing Instruments to Assess Residency Applicants

Glen Bandiera, MD, Med, FRCPC

Implicit in the considerable effort programs spend assessing applicants is the belief that suitability for the program and future physician performance can be predicted. For this to be true, programs must start with a reliable and valid applicant assessment protocol. We have developed a structured application assessment instrument¹ as well as a structured interview scoring instrument² and evaluated them in our emergency medicine program. We believe that program improvement efforts must include attention to admission of residents and offer a description of our approach to developing and evaluating these instruments as a practical example for other program directors to consider.

What are the characteristics of good evaluation instruments?

Residency program committees assessing applicants face fiscal, scheduling and other practical constraints. The imperative is thus to design evaluation instruments that maximize reliability and validity while allowing the most efficient overall process. Reliability is the degree to which variability among applicants' scores is due to variations between the applicants themselves ('true variance', or 'signal') and not to variation in confounding factors such as who does the rating, when the rating occurs, or the order in which candidates are assessed ('false variance', or 'noise'). A reliable instrument could also be considered valid if it assesses aspects of a candidate believed to be important (construct validity), produces results correlated with other objective measures of performance (criterion validity), predicts future performance (predictive validity), or has credibility in the eyes of those involved (face validity).

While variability arising from the use of different evaluators could be overcome with a single person scoring each package or interview, this is usually impractical and most program directors would be reluctant to base such high stakes decisions on a single person. Most admissions will therefore be done by committee, with each candidate being assessed by one of several teams. This necessitates instruments that minimize variability in scores due to the use of multiple evaluators. An average of several evaluators' scores is likely to be more reliable and stable than a score from a single evaluator. A well-designed instrument can minimize the number of evaluators per applicant needed to achieve a stable average score and more importantly, obviate the need for the same set of assessors to review every applicant.

From a reliability perspective, the organization of teams is not important as long as multiple evaluators are used and decisions are made on average or aggregate scores. For example, two teams of three interviewers can generate scores

as reliably as three teams of two. We have used such instruments in our Emergency Medicine residency program for several years and have found them to generate highly reliable scores. We need to average at least two evaluators' scores for each applicant and do not experience inadequate reliability despite using multiple teams.

Instruments must be tailored to both the program's needs and the materials available from eligible candidates. Despite a lack of conclusive evidence that information obtained during candidate assessment reliably predicts future performance, most program directors rely on a broad based application and interview assessment protocol. Self-assessments, previous experiences, results from objective knowledge tests, stated career objectives and references are commonly included in application packages. Program directors report heavy reliance on interviews, during which interpersonal skills, problem-solving ability, responsibility, initiative, self-assessment, communication and teamwork are commonly assessed. Admissions committees need to decide which candidate attributes to measure, where in the assessment protocol to evaluate each, and how best to score each candidate's merit with respect to each attribute. Instruments must then be designed for each step in the process, most commonly resulting in separate instruments for application packages and for interviews.

Fundamental principles guide the development of reliable instruments regardless of their intended use. Asking for judgments on a multitude of categories invites "halo error", meaning that assessments of specific traits are influenced by an overall impression of a candidate. Two common methods used to overcome this limitation are 1) making scoring criteria very explicit and 2) limiting the number of independent attributes to which a score will be assigned. Most instruments solicit either a numeric score or use a rating scale such as the visual analog scale (in which an 'x' is placed on a continuous line) or Likert scale (classically using discreet options from 'strongly disagree' to 'strongly agree'). While neither method is clearly superior, both should be anchored with specific descriptors detailing what level of performance is to be associated with a given score. For example, candidates may be asked about positions of leadership. Scoring scales should include statements such as:

-
1. Could not identify a leadership experience
 2. Could identify but not describe a leadership experience
 3. Could describe a leadership experience but not personal development from it
 4. Could demonstrate personal growth from a leadership experience
 5. Multiple examples of personal growth or excellence in leadership

Such a scale will be more reliable than one based on such abstract terms as 'poor' and 'good'. Each scale should have different descriptors, tailored to the applicable attribute. Finally, scales should have a defined middle option and two

feasible extremes. This will maximize use of the entire scale, improve discriminative ability, and avoid ceiling/floor (or 'hawk'/ 'dove') effects.

It is generally felt that people cannot independently assess more than three or four aspects of any given experience. This has been shown to be true in multiple scenarios such as clinical rotations, oral examinations, interviews, and application package assessments. An ideal instrument, therefore, would ask evaluators to provide 3 or 4 independent scores per applicant using detailed descriptors for each rating scale. Instruments should ideally be pilot tested and revised prior to use. Thus designed, an instrument should deliver reasonable reliability and allow a minimum number of evaluators.

An application assessment instrument

Our application assessment instrument uses a series of numeric scores with clearly defined anchors for each of four application components.¹ Our interview scoring instrument has evolved from a numeric scale to a series of modified Likert scales, allowing a less structured interview process.² Interviewers are also asked for a global assessment of each candidate.

For each resident application package, independent scores can be solicited for attributes such as 'leadership' and 'social responsibility' or alternatively for different components of the package such as 'curriculum vitae' or

“We prefer candidates who excel in all four components of the application package and have therefore set minimum performance standards for each component. Previously, candidates achieving the minimum score in each area were granted an interview, irrespective of their overall score. Dichotomizing analog scores in this manner results in a decrease in reliability because information is lost.”

'personal statement'. While our instrument uses the latter method, it is not clear which is superior. The former allows the incorporation of various sources of information into the assigned scores (for example, evidence of 'leadership' might appear on a CV and in a personal statement). The latter allows for more concrete descriptors because they can be tailored to the format and type of information (e.g., a CV score might include 'elective experience', 'work experience', 'volunteer activity', and 'awards', each with its own score).

Evaluators using our format have expressed concern that information contained in one component of the application package can be vital to clarifying or embellishing information

in another. Unless program committees using this method are prepared to insist that component scores be assigned based only on information contained in that component or accept that component scores are not independent, they may find assigning scores to candidates' attributes rather than application components a more acceptable approach.

Regardless of approach, each score can itself be calculated from a number of related sub-scores or simply be guided by a list of typical examples of what to look for. For example, in our instrument, assessors are asked to calculate the CV score by first determining the number of hours/days/weeks for various experiences and assigning fixed numbers to certain accomplishments such as winning a scholarship or completing a postgraduate degree. The CV score is the sum of these numbers. Alternatively, an assessor might be asked to score a CV using a ten point scale for the following items: breadth of experience in emergency medicine; evidence of initiative in research, project design or curriculum development; social responsibility defined by volunteer work or advocacy; and leadership defined by political, business, sports, or academic responsibilities. A similar approach would apply to other application components.

An interview scoring instrument

Interview scoring instruments are designed in a similar manner. Once attributes of interest and specific descriptions for each level of performance have been defined, an interview schedule must be developed. Each interview team should be asked to assess in depth only one or two major candidate attributes, and at least two interviewers should score each attribute. This helps to reduce halo effect, increase independence of attribute scores, and focus the interviewers' questions. Candidates' interview scores are the average of all interviewers' scores.

Using the application and interview data for admission decisions

Candidates will eventually need to be ranked. Some program committees may use application scores to screen candidates prior to interviews and others may interview all applicants. In addition, final ranks may be based on either a combined application and interview score or entirely on the interview score. Both strategies are justifiable and choice of strategy does not diminish the importance of rigorous instrument design. We prefer candidates who excel in all four components of the application package and have therefore set minimum performance standards for each component. Previously, candidates achieving the minimum score in each area were granted an interview, irrespective of their overall score. Dichotomizing analog scores in this manner results in a decrease in reliability because information is lost. We have since changed our approach to establish a fixed number of interview spots and interview candidates with the top overall scores who also achieve the minimum required score in the

four application components. We base our ranks on combined interview and application scores. Our experience with assessment instruments has lead to average score reliabilities up to .90 with two assessors.^{1,2}

Do these admissions instruments predict future performance?

Our instruments have high face validity because they involve concrete information related to program objectives. Predictive validity, however, has been difficult to establish. Some, but not all, evidence suggests that performance on objective measures of knowledge can predict future performance on similar knowledge tests. There is little evidence, however, that the remaining data incorporated into admissions decisions correlates well with future performance. This may result from the multi-factorial nature of professional competence and the lack of precise instruments for its measurement, rather than a true lack of relationship between admissions scores and future performance. Medical residents, having succeeded in many prerequisites, are a homogeneous group and the failure rate for residents is uniformly low. This low variability makes demonstration of correlation more difficult. For example, we have an average of 50 applicants for 3 residency positions per year and have the luxury of admitting uniformly strong candidates. Regrettably, this small number of deserving candidates and the imprecision of in-training assessments make it difficult to prove the predictive validity of our instruments. Residents in our program have thus far met expectations generated during the admissions process.

Program committees have a responsibility to adhere to an admissions protocol that provides an accurate, fair appraisal of applicants. Only by attending to this process can we assure ourselves that we are indeed selecting the most appropriate candidates for our programs and move on to tackling the important task of assuring that our residents become competent physicians able to practice independently. ■

Glen Bandiera, MD, MEd, FRCPC is Program Director for Emergency Medicine and Director of Postgraduate Medical Education at St. Michael's Hospital, and Assistant Professor, Division of Emergency Medicine, University of Toronto in Ontario, Canada.

¹ Bandiera G, Regehr G. A structured application assessment instrument is a reliable instrument for assessing applications to Canadian postgraduate training programs in emergency medicine. *Acad. Emerg. Med.* 2003; 10(6): 594-598.

² Bandiera GW, Regehr, G. Reliability of a structured interview scoring instrument for a Canadian postgraduate emergency medicine training program. *Acad. Emerg. Med.* 2004; 11(1): 27-32.

A Postgraduate Orientation Assessment at the University of Michigan

*Jennifer C. Janus MD, Stanley J. Hamstra, PhD,
Lisa Colletti, MD, Monica L. Lypson, MD*

Medical students' clinical skills often vary at the start of their first year of postgraduate medical education, as has been demonstrated in several studies.^{1,2,3} Furthermore, there is a perception of need for objective evaluation of residents' clinical skills at the start of their residency, as demonstrated in a study by Langdale and colleagues. In her survey of program directors from five different residency programs, she reported: "Virtually all the residency directors indicated that they expected to spend a portion of the first three to six months of residency confirming competence in key skills and competencies."⁴

Beginning in 2002, the University of Michigan developed an incoming first-year resident Objective Standardized Clinical Examination (OSCE), named the Postgraduate Orientation Assessment (POA).⁵ The POA focuses on all six of the ACGME/ABMS general competencies. It is an 8–10 station

"Topics were chosen with a focus on skills likely to be needed by residents in the first four months of their training. These stations utilize various forms of education and assessment, including standardized patients, PowerPoint presentations, videos and Web based format."

OSCE delivered during orientation and includes: informed consent, aseptic technique, cultural communication, evidence-based medicine, pain assessment, systems compliance (fire safety/restraint policy), geriatric functional assessment, patient safety (order writing/respiratory distress), radiograph interpretation, and critical lab values. Topics were chosen with a focus on skills likely to be needed by residents in the first four months of their training. These stations utilize various forms of education and assessment, including standardized patients, PowerPoint® presentations, videos and Web based format. Feedback/remediation from each station, usually in the form of take-home literature, is provided at the time of the assessment. A complete description of each station is described in Lypson et. al.⁵

The average overall score for the first administration of the POA in 2002 (132 first-year residents from 14 different specialties and 59 different schools) was 74.8%, with a range of 58.6–90.0%. Construct validity was examined by comparing these results with the performance of the University of

Michigan Medical School (UMMS) graduates, since some of the stations were based on an OSCE administered to the medical students. Indeed, UMMS graduates performed better on the informed consent, evidence-based medicine, and cross-cultural communication stations, even when scores were adjusted for performance on the USMLE.⁵

Evolution of the POA

Over the past several years, there have been slight modifications to the assessment. The changes often reflect important content areas that program directors think are missing or those that the institution sees a need to add. We have tended to focus on Joint Commission on Accreditation of Hospital Organizations (JCAHO) national patient safety goals and how they affect the care of our patients. We theorize that first-year residents have no other exposure to these important areas of hospital function outside of the POA. For instance, the evidence-based medicine station now tests for the ability to identify abstracts relevant to two clinical questions, where formerly it tested just one. Questions were added to the critical values station to cover medication errors. The aseptic technique station now uses standardized patients, as does the Universal Protocol for Preventing Wrong Site, Wrong Procedure and Wrong Person Surgery™.⁶ The pain assessment station now tests for the ability to calculate equianalgesic dosing.

Two new stations have been added to the POA. The patient safety/order writing station (formalized in 2004) is a multi-step, paper-and-pencil simulation of an asthma patient in the emergency department. In addition to testing for knowledge regarding the treatment of an acute asthma exacerbation, its goal is to assess: 1) the ability of residents to recognize impending respiratory failure and to ask for help when appropriate, and 2) proper order writing technique. Core competencies tested include patient care, medical knowledge, systems-based practice, and professionalism. The average score on this station for incoming PGY1 residents was 82.6% (SD 11.3%) in 2004 and 75.8% (SD 9.9%) in 2005. A large minority of residents failed to ask for help as the simulation prompted the patient's status to decline. Remediation was

“Remediation was provided in the form of the University of Michigan Hospital System clinical guidelines for the treatment of asthma, and a card describing the National Patient Safety Goals.”

provided in the form of the University of Michigan Hospital System clinical guidelines for the treatment of asthma, and a card describing the National Patient Safety Goals. These results prompted a recommendation to program directors to reinforce the need for legibly written, dated, timed and signed orders and when to ask for assistance from senior residents, nurses and attendings.

In 2005 we added a geriatric functional assessment station. Using a standardized patient, this station tests residents' ability to perform assessment of functional status and affect, and gather appropriate data from the history including falls, affective disorders and incontinence. Residents are also expected to screen for cognitive impairment using what tools they might have in their armamentarium.⁷ This station evaluates the core competencies of patient care, medical knowledge, systems-based practice, professionalism, and interpersonal and communication skills. Given lack of geriatric curriculum in most medical schools, the average score for incoming first-year residents (68%, SD 10%) was not unexpected. Remediation is provided in the form of handouts. This station was particularly well received by the residents, with 56% of residents agreeing or strongly agreeing that “this station enhanced my ability to evaluate the functional status of an older patient” and 63% strongly agreeing or agreeing that “this station will motivate me to spend additional time learning how to assess older patients”.

Reflections and Prospects

The POA is currently a ten-station, three and one half hour assessment (including breaks). In the first year, participation was optional at the discretion of the program director. From 2002 forward it has been required of all incoming first-year residents during orientation. A total of 570 new residents from 15 programs (Anesthesiology, Emergency Medicine, Family Medicine, Internal Medicine, Internal Medicine / Pediatrics, Obstetrics & Gynecology, Pathology, Pediatrics, Psychiatry, Surgery, Surgery-Prelim, Plastic Surgery, Neurosurgery, Surgery-OMF, Orthopedics, Urology) have been evaluated during orientation (2002-2005). The POA has been well received over the past four years; an average of 79% of residents feel that the POA is a useful way to spend orientation. In four administrations on average 78% of first-year residents recommend the POA for new residents in the following year. More importantly, on average 74% stated they had learned a new clinical skill during the assessment. The POA has since received a commendation during institutional accreditation review.

Looking forward, the Graduate Medical Education Committee is currently interviewing residency program directors to assess the usefulness of the data generated by the POA, as well as gathering suggestions for change. To date, fourteen of nineteen program directors have responded (74%). Of respondents, 100% state that they review the data, and a large majority (86%) feels that the data reflects situations likely to arise in residency. Three program directors stated that they implemented remediation for individuals based on POA scores. A few programs implemented program-wide education based on POA results, specifically in the areas of informed consent and aseptic technique. Eleven program directors rated the usefulness of the data on a scale of 1 to 5 (with 5 being extremely useful), with a mean score was 3.32 (range 2–4).

Despite multiple attempts to identify more convenient times, orientation continues to be the best time to have all first-year residents participate in this assessment. Beyond the

Practice-Pattern-Based Curriculum Development

Richard S. Isaacson, MD

practical requirements of orientation, the advantages of using this time period to emphasize certain core aspects of clinical care and hands-on education far out-weigh the benefits of later testing. The estimated cost continues to hover around \$225-250 per resident (\$500 per resident, if additional salary dollars are factored in). Thus, it continues to be costly in its implementation and may prove to be cost prohibitive to other programs. Nevertheless, aspects of the POA can be transported to other institutions. Interactive formats (e.g., web based) can be implemented in all programs, changing “hospital orientation” from a “bore” to an informative exercise that no one should miss.

The POA has been a useful tool for documenting the baseline competency of incoming residents, is broadly applicable to all specialties, and can be modified to meet the changing needs in medical education for excellent patient care and safety. We have seen program directors begin to use POA data to modify their programs by targeting areas of weakness and creating individualized learning plans for specific residents. It has the potential for being used to track resident progress, as well as document program effectiveness (for example by aggregating PGY1 scores for a particular program and comparing them to the aggregated scores of the same residents at the end of their training). With phase 3 of the Outcome Project scheduled to begin July, 2006, programs need to provide evidence that they are making data-driven improvements. The POA offers a possible approach that could help some program directors and institutions meet this requirement. ■

Jennifer C. Janus, MD, is a 4th year Medicine/Pediatric Resident and was part of the first resident cohort to take the examination; Stanley J. Hamstra, PhD, is an Associate Professor of Medical Education and leads the UM residency education efforts; Lisa Colletti, MD, is Associate Dean of Graduate Medical Education and Professor of Surgery; Monica L. Lypson, MD, is Assistant Dean of Graduate Medical Education and Assistant Professor of Internal Medicine, all are at the University of Michigan Medical School, Ann Arbor, MI.

Curriculum reform that positively influences resident performance outcomes is the essential aim of the ACGME’s Outcome Project. We have implemented an innovative curriculum development model, based on the study of practice patterns, which meets the ACGME accreditation requirement (effective July, 2006) for programs to demonstrate that they are making data-driven improvements. We describe how this model was used for program improvement in three specialty areas at two academic medical centers (Mount Sinai Medical Center/University of Miami, Miami Beach, FL; Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA).

Criteria used for curriculum development

Curricula are commonly “developed” almost entirely through expert opinion rather than a rigorous scientific and reproducible process. An evidence-based process is essential toward optimizing what we teach to physicians-in-training.

“Curricula should adapt to the needs of the learner. For example, in anticipation of the influenza season, it makes sense to focus teachings on the epidemiology, diagnosis, treatment and strategies to prevent the spread of the disease.”

Curricula should be **representative of both high-yield topics and uncommon presentations** that residents and fellows encounter on an every day basis. Frequent diagnoses can be determined by examining a database of the most common admitting diagnoses. The learning needs of the student should be included, and one mechanism for this is to survey residents to **identify areas of least familiarity**. Curricula should **adapt to the needs of the learner**. For example, in anticipation of the influenza season, it makes sense to focus teachings on the epidemiology, diagnosis, treatment and strategies to prevent the spread of the disease. Also, if during one week there is a flurry of patients admitted with atrial fibrillation, it makes sense to offer an adaptable curriculum which can focus on that topic during that week. A **constant emphasis on health system issues** must also be incorporated. For example, the rising costs of health care in America create an imperative to focus on utilization issues, which is the hallmark of the competency of systems-based practice.

- ¹ Burch VC, Nash RC, Zabow T, Gibbs T, Aubin L, Jacobs B, and Hift RJ. A structured assessment of newly qualified medical graduates. *Med Ed.* 2005; 39: 723-731.
- ² Bahman J and Devries JM. Evaluation of clinical competence: the gap between expectation and performance. *Pediatrics.* 1996; 97: 179-184.
- ³ Moercke AM and Eika B. What are the clinical skills levels of newly graduated physician? Self-assessment study of an intended curriculum identified by a Delphi process. *Med Ed.* 2002; 36: 472-478.
- ⁴ Langdale LA, Schaad D, Wipf J, Marshall S, Vontver L, and Scott CS. Preparing graduates for the first year of residency: are medical schools meeting the need? *Acad Med.* 2003; 78: 39-44.
- ⁵ Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO. Assessing residents’ competencies at baseline: identifying the gaps. *Acad Med.* 2004; 79: 564-570.
- ⁶ Joint Commission on Accreditation of Healthcare Organizations. Universal Protocol for Preventing Wrong Site, Wrong Procedure, Wrong Person Surgery™. Available at http://www.jcaho.org/NR/rdonlyres/E3C600EB-043B-4E86-B04E-CA4A89AD5433/0/universal_protocol.pdf Accessed March 7, 2006.
- ⁷ Williams BC, Hall KE, Supiano MA, Fitzgerald JT, Halter JB. Development of a Standardized Patient Instructor to Teach Functional Assessment and Communication Skills to Medical Students and House Officers. Submitted for Publication.

We show in the following examples that focused teaching interventions can improve both resident performance outcomes and appropriate utilization of costly hospital resources. A favorable cost/benefit analysis provides financial incentive for improving medical education, and this observation has been an exciting realization for both hospital administrators and medical educators alike.

Three examples of practice-pattern curriculum development, implementation and outcomes

Internal Medicine

Our first step in practice-pattern based curriculum development for internal medicine was to identify and characterize the most common diagnoses that residents and fellows encounter during their training. We retrospectively reviewed over 15,000 consecutive patient admissions to the

“The curriculum was implemented through a lecture series focused on the most common primary IM admitting diagnoses. We prospectively collected daily IM admitting diagnoses and had residents record which diagnoses they wanted to learn more about.”

Internal Medicine (IM) resident teaching service of Mount Sinai Medical Center, a large urban teaching hospital. Over the last 10 years, admitting diagnoses have been collected through daily intern admission sheets which record admitting diagnosis, as well as daily resident duty hours. These diagnoses were entered into Microsoft FoxPro® by the IM Program Coordinator and then manually counted over a long weekend. The primary outcome measure was most common admitting diagnoses, and a preliminary outline for a comprehensive core curriculum for IM residents was developed based on this data.

Applying the latest in evidence-based medicine and national performance improvement initiatives, this outline was then expanded by incorporating recommendations of the American College of Physicians Clinical Practice Guidelines, the National Guidelines Clearinghouse, US Preventive Services Task Force recommendations, Centers for Medicare and Medicaid Services (CMS) Hospital Quality Initiatives (HQI) program and the CMS Physician Voluntary Reporting Program (PVRP) measures.¹⁻⁵ The curriculum was finalized after review by the IM Education committee, which is a committee appointed by physicians in Mount Sinai’s Research Unit in Medical Education (RUME). The curriculum was implemented through a lecture series focused on the most common primary IM admitting diagnoses. We prospectively collected daily IM admitting diagnoses and had residents record which diagnoses they wanted to learn more about. This was achieved by adding a simple “check-box” to the daily admitting diagnosis sheets.

A bi-weekly, adaptable “rotating curriculum” based on the most common admitting diagnoses of the previous week or areas of least familiarity as suggested by daily resident data collection was subsequently implemented.

The twelve most common primary IM admitting diagnoses were: chest pain/rule out myocardial infarction, stroke/transient ischemic attack, pneumonia, syncope/near syncope, gastrointestinal bleed, congestive heart failure, fever, abdominal pain, cellulitis, COPD, altered mental status, and atrial fibrillation.⁶ For each of these topics, we selected recent articles from the literature and used this as an adjunct to our teaching. For example, for congestive heart failure, we used the ACC/AHA guideline update for the diagnosis and management of chronic heart failure in the adult.⁷ Resident-requested topics of additional teaching have been wide in range (e.g., follicular thyroid carcinoma, osteonecrosis of the knee, conversion disorder) and these topics have added a unique and flexible aspect to resident didactics.

This practice-pattern and “Quality Agenda” competency-based curriculum development model addresses the most common diagnoses encountered by our IM residents and their self-designated areas of least familiarity. Pre-and post-focused teaching intervention testing with emphasis on knowledge of national consensus guidelines will begin in July 2006, in accordance with the ACGME accreditation requirement.

Neurology/Radiology

To address the issue of costly over-utilization of radiographic images ordered from the Emergency Room (ER), we sought to determine if an evidence-based focused teaching intervention would improve clinical practice.

Low back and neck complaints are common in the ER and while guidelines exist for diagnostic evaluation, management remains a challenge and costly over-utilization of neuroimaging is prevalent. We developed a comprehensive curriculum for Emergency Medicine (EM) residents using the practice-pattern based database model and then assessed whether alterations in clinical practice occurred after implementation.

Over a five-month period, we collected a series of 1,100 Neurology consults requested from the ER at Beth Israel Deaconess Medical Center.⁸ The database included patient demographics, reason for consult, differential diagnosis made by the EM resident, neuroimaging study ordered, and final diagnosis reached by neurology consultant. Outcome measures included most common Neurology consultations requested, comparison of differential diagnosis of the EM resident to final diagnosis of the Neurologist, and necessity of neuroimaging study ordered. The five most common Neurology consults were: Stroke/Transient Ischemic Attack, Seizure, Neck/Back pain r/o Spinal cord compression, Dizziness/vertigo, and Headache. These five diagnoses comprised >60% of all consults called and this became the foundation for a new comprehensive evidence-based Neurology core curriculum for EM residents, utilizing policy statements of the American Academy of Neurology (AAN) and the American College of Emergency Physicians (ACEP).⁹⁻¹²

Table 1

Alterations of Practice Patterns in the ER**Frequency of Back/Neck Consults**

Pre-intervention 11.6% (128/1100)	Post-intervention (1 month) 10.2% (27/266) (Chi-square=.47, $p \leq 1$, NS)	Post intervention (3 months) 10.4% (82/788) (Chi-square=.70, $p \leq 1$, NS)
--------------------------------------	--	---

Consults

Ave # consults/mo 5 mo pre-intervention 30.2	Ave # consults/mo 3 mo post-intervention 27.3
---	--

% Unnecessary Neuro-Imaging Studies Ordered by EM Resident (per consult)

Pre-intervention 16.1%	Post-intervention (1 month) 12.4% (Chi-square 2.24, $p \leq .20$)	Post-intervention (3 months) 13.4% (Chi-square 2.54, $p \leq .20$)
---------------------------	--	---

Cost Savings

Estimated cost due to unnecessary studies \$43,600/100 consults	Post-intervention cost savings ~\$10,000/100 consults	Projected cost savings ~\$34,000/year
--	--	--

We found that the most prevalent over-utilization of neuroimaging studies was related to the back/neck pain consults. We thus developed a curriculum based on this primary outcome measure and refined the curriculum by incorporating consensus recommendations of the AAN.¹³ The curriculum was finalized after review by Department Chiefs of Neurology (Dr. Clifford B. Saper), Neuroradiology (Dr. David Hackney), and Emergency Medicine (Dr. Jonathan A. Edlow), and Director of Clinical Services in Neurology (Dr. Michael Ronthal) and was implemented through didactics and through the adoption of a hospital policy statement on how to appropriately evaluate and treat these patients. Didactics included history and physical exam for back/neck complaints, management issues, guidelines for neuroimaging performance improvement, and handouts of summary pocket cards (e.g., indications for early/immediate MR imaging, when to use and not to use contrast).

As shown in **Table 1**, we found that one-month following curriculum implementation, the frequency of back/neck consults decreased by 12.5%. Unnecessary or incorrectly ordered neuroimaging studies decreased by 23% to 12.4% per consult. Three months following curriculum intervention, the frequency of back/neck consults decreased by 10.3%, and unnecessary or incorrectly ordered neuroimaging studies decreased by 18.2% per consult. While a larger sample size will be necessary to achieve statistical significance, these findings suggest that improved consults and accuracy of imaging modality ordered is possible with a focused EM core curriculum. In addition, the cost/benefit analysis showed favourable results.¹⁴

Surgery

In order to improve adherence to evidence-based practice for the use of medical consults in surgical subspecialties, we used

Table 2

Common Consultations and Management Differences**Most common medical consultations requested**

Medical clearance prior to surgery	23%
Management of Hypertension	16%
Management of Diabetes	13%
Management of Obstructive Lung disease	7%
Management of IHD, CHF and Cardiac arrhythmias	7%
Anticoagulation	7%
Chest pain	6%

Most common management differences between medical and surgical services

Antibiotic prophylaxis	14%
VTE prophylaxis indications	14%
Length of antibiotic therapy	10%
Transfusion	3%
Nutrition, TPN	3%
Other (e.g., pain control, end of life care, blood pressure mgmt)	5%
None Identified	51%

Table 3

Increased Resident Knowledge and Satisfaction with Curriculum**Medical Clearance prior to surgery Pre- and Post- test results for IM Residents**

Pre-test results: 60% correct (n=27)
 Post-test results: 80.8% correct (n=24)

Antibiotic prophylaxis Pre- and Post-intervention results for surgical residents

Pre-test results: 68.9% correct (n=14)
 Post-test results: 84.6% correct (n=13)

Survey of IM resident satisfaction, based on 5 point Likert scale (n=24)

Implemented curriculum	4.7
Will improve clinical practice	4.8
Useful for boards preparation	4.2

Survey of surgery resident satisfaction, based on 5 point Likert scale (n=13)

Implemented curriculum	4.4
Will improve clinical practice	4.1
Useful for ABSITE/boards preparation	3.2

the practice pattern-based improvement model and national consensus guidelines to develop curricula for surgical and IM residents. We assessed knowledge both pre- and post intervention, as well as resident perceptions.

With colleagues Antonio C. Cano, MD, and Tariq Al-Musawi, MD, we prospectively collected medical consult patient information including demographics, admitting diagnoses, reasons for medical consult, post-operative complications, length of hospitalization (including rehabilitation), resident-rated case complexity (scale 1 to 7), and physician-to-physician verbal interactions when management preferences differed between surgical and medical

teams, stratifying outcomes into most common consults requested and management preference differences (**Table 2**).¹⁵

Core curricula for surgical and medical residents were developed based on the most common consult questions and national consensus guidelines, including those from the American Heart Association for peri-operative evaluation, American College of Chest Physicians for venous thromboembolism prophylaxis, American Society of Infectious Disease for antibiotic prophylaxis, and literature-based expert opinion for the prevention and management of post-operative pulmonary complications, pain control, nutrition, bleeding disorders, and the use of insulin and steroids in the peri-operative period.¹⁶⁻¹⁸

Table 4

Ongoing Curriculum Enhancement Projects

Specialty	Title of Project
Geriatrics (Dementia)	Developing a dementia curriculum for IM residents (<i>Amiel Levin MD</i>)
Internal Medicine	Practice-pattern and evidence-based curriculum development for IM residents (<i>Gloria Weinberg MD, Allen Young MD</i>)
Sleep Medicine	How many polysomnograms must sleep fellows score before becoming proficient at scoring sleep? (<i>Alejandro Chediak MD</i>)
Sleep Medicine	Developing a practice-pattern based Sleep Medicine curriculum for IM residents
Medical Education/Neurology	Attitudes of Osteopathic vs. Allopathic Medical Students toward Adult Neurology Training (<i>Shara Steiner, MS-3</i>)
Urology	Men's Health Initiative – Practice-pattern based Urology curriculum development for IM residents (<i>Alan Nieder MD</i>)
Radiology/Neurology	Diagnosis of back and neck complaints in the ER: Does focused teaching intervention improve clinical practice? (<i>R. Isaacson MD</i>)
Infectious Disease	Developing a practice-pattern based ID curriculum for IM residents (<i>Kenneth Ratzan MD</i>)
Surgery	Standardizing healthcare delivery for surgical patients (<i>Antonio C. Cano MD, Tariq Al-Musawi MD</i>)
Neurology	American Academy of Neurology – Practice-pattern based Medical Student Clerkship Curriculum Development (<i>Richard S. Isaacson MD</i>)

For further information on these projects, including examples of specialty-specific practice pattern-based curricula, and pre- and post- testing and evaluation tools, please contact Richard Isaacson at risaacso@msmc.com or visit <http://www.knowledgebasedlearning.org/curriculum.html>.

We developed a curriculum for surgical risk stratification/medical clearance prior to surgery for IM Residents, testing knowledge about national consensus guidelines both before and after implementation. For residents that were unable to attend lectures due to scheduling conflicts, didactics were digitally recorded for viewing at a later date. We implemented the developed curriculum for antibiotic prophylaxis in the same manner for surgical residents, testing their knowledge both before and after implementation. We also distributed summary pocket cards of the didactics which residents could keep with them in their lab coats for quick reference.

“Both internal medicine and surgery residents demonstrated increased knowledge and satisfaction with the curriculum.”

Both internal medicine and surgery residents demonstrated increased knowledge and satisfaction with the curriculum (**Table 3**). While medicine residents felt the curriculum was useful for board preparation (4.2/5), surgery residents did not feel their curriculum was as useful (3.2/5).

We have implemented a validity study and cost/benefit analysis to assess the effectiveness of these curricula on practice patterns. Through collaboration with the Director of Pharmacy at Mount Sinai Medical Center, we have instituted a pilot program to track surgical resident antimicrobial use before and after our focused teaching intervention, comparing practice-patterns of residents who participated in the educational intervention to those residents not in attendance. We are also tracking ordering of antibiotics used for prophylaxis and identifying which resident ordered the agent, and will compare timing of peri-procedure administration and duration of antibiotic use following IRB approval for record retrieval. An additional benefit of this project has been the opportunity for medical students and residents to be involved with outcomes research in medical education.

Conclusions

While patient management differences are inherent to the practice of medicine, an evidence-based approach utilizing national consensus guidelines is a mechanism for both improving resident education and standardizing health care delivery in academic medical centers. Our practice pattern-based model may be used as a generalizable tool to enhance curriculum development in all specialties and also serve as a mechanism for outcomes assessment of the ACGME core competencies. To build and expand on this model, a new division within the Department of Medical Education at Mount Sinai Medical Center, the Research Unit in Medical Education (RUME), was established in July 2005. Ideas for outcomes projects are discussed at monthly Graduate Medical Education Committee meetings, and RUME staff facilitates research by interested faculty and medical educators from a variety of specialties. Several projects are planned or underway, shown in **Table 4**. ■

Richard S. Isaacson, MD, is Director of the Research Unit in Medical Education, and Associate Medical Director of the Wien Center for Alzheimer's Disease and Memory Disorders at Mount Sinai Medical Center/University of Miami

- 1 American College of Physicians. Clinical Practice Guidelines. 15 October 2004: <http://www.acponline.org/sci-policy/guidelines/> (accessed January 31, 2006)
- 2 Agency for Healthcare Research and Quality. National Guideline Clearinghouse. 31 January 2006: <http://www.guideline.gov/> (accessed January 31, 2006)
- 3 Agency for Healthcare Research and Quality. U.S. Preventative Services Task Force: <http://www.ahrq.gov/clinic/uspstfix.htm> (accessed January 31, 2006)
- 4 United States Department of Health and Human Services. Hospital Compare: Hospital Quality Measures. 1 September 2005: <http://www.hospitalcompare.hhs.gov/> (accessed January 31, 2006)
- 5 Centers for Medicare and Medicaid Services. Physician Focused Quality Initiative: Physician Voluntary Reporting Program. 9 February 2006: <http://www.cms.hhs.gov/PhysicianFocusedQualInits/> (accessed February 13, 2006)
- 6 Isaacson RS, Young DA, Weinberg G. Educational Innovations Project Pilot Project: Developing a core and rotating Internal Medicine Curriculum. http://www.acgme.org/outcome/conferences/pbli_abstractP13.pdf (accessed February 13, 2006)
- 7 Hunt SA et al. ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Bethesda (MD): American College of Cardiology Foundation (ACCF); 2005 Aug. 82 p.
- 8 Isaacson RS, Schussler E, Edlow J. Developing a neurology curriculum for emergency medicine residents. *Neurology* 2005; 64 (suppl):A9.
- 9 Adams H, Adams R, Del Zoppo G, Goldstein LB. AHA/ASA Scientific Statement: Guidelines for the early management of patients with ischemic stroke. *Stroke* 2005; 36:916-921.
- 10 Practice Parameter: Neuroimaging in the emergency patient presenting with seizure. *Neurology* 1996; 47:288-291.
- 11 ACEP Clinical policy: critical issues in the evaluation and management of patients presenting to the emergency department with acute headache. *Ann Emer Med* 2002; 39:108-122.
- 12 Fife TD, Tusa RJ, Furman M et al. Assessment: Vestibular testing techniques in adults and children: Report of the Therapeutics and Technology Assessment Subcommittee of the AAN. *Neurology* 2000; 55:1431-1441.
- 13 Practice Parameters: Magnetic resonance imaging in the evaluation of low back syndrome (Summary statement). *Neurology* 1994; 44:767-70.
- 14 Isaacson RS, Schussler E, Ronthal M, Edlow J. Diagnosis and management of neck and back complaints in the ER: Does focused teaching intervention improve clinical practice? http://www.acgme.org/outcome/conferences/pbli_abstractP14.pdf (accessed February 13, 2006)
- 15 Cano-Gomez A, Isaacson RS, Weinberg G. Standardizing healthcare delivery of surgical patients. http://www.acgme.org/outcome/conferences/pbli_abstractP4.pdf (accessed February 13, 2006)
- 16 Bratzler DW, Houck PM. Antimicrobial Prophylaxis for Surgery: An Advisory Statement from the National Surgical Infection Prevention Project. *Clinical Infectious Diseases* 2004; 38:1706-1715.
- 17 ACC/AHA Guideline Update for Perioperative Cardiovascular Evaluation for Noncardiac Surgery. *J Am Coll Cardiol* 2002; 39:542-53.
- 18 Hirsh J, Guyatt G, Albers G, et al. The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy: Evidence-based guidelines. *Chest* 2004; 126(suppl):1725-1735.

Journal Club: A Tool to Teach, Assess, and Improve Resident Competence in Practice-based Learning and Improvement

Andrew G. Lee, MD

The ACGME-mandated competency, practice-based learning and improvement (PBLI), is particularly difficult for programs because traditional methods for teaching and assessing residents (e.g., lectures, role-modeling, and resident rotations; and global evaluations and written or oral exams) are poorly designed to measure the actual ability to acquire, synthesize, apply, and sustain over time evidence-based medicine (EBM) in a real world clinical setting. The University of Iowa's Department of Ophthalmology approached this challenge by forming an internal task force on the ACGME competencies to develop and test tools that could be adapted for use within our existing curriculum.

Table 1
"Best Practices" for a Successful Journal Club⁴

- Mandatory attendance
- Positive financial and philosophical support for the journal club from the department, the chair, and the program director
- Designated, committed, experienced, and skilled moderators (e.g., faculty supervisor)
- Formal, consistent, and regular schedule and location (e.g. monthly with same location, stable attendance)
- Regular provision of food and drink
- Protected and convenient time (e.g., evening journal club)
- Structured checklist and explicitly defined written goals
- Limited number of articles but reviewed in more depth and articles selected and distributed with sufficient time (weeks) to read completely
- Emphasis on original research articles (e.g., clinical trials)
- Inclusion of basic but formal epidemiology and statistics and principles of evidence-based medicine
- Resident journal club independent of faculty journal club
- Incorporation of adult learning principles (e.g., rationale for study, self-directed learning, application of reservoir of learner experiences, task-centered approach, emphasis on practical application of information, emphasis on problem solving skills)
- Reinforcement of critical information with active instructional feedback
- Active participation (rather than simply attendance) with interactive discussion format (e.g., exchange of ideas on controversial topics)
- Written documentation of participation and completion of checklist

Table 2
Goals for a Standardized Journal Club⁴

- Acquire, disseminate, and apply new medical information
- Teach and assess critical appraisal skills for reading and writing a scientific paper
- Promote life long learning skills in evidence-based medicine
- Improve reading habits
- Provide an interactive and social opportunity for peer-to-peer learning
- Improve small group participation, presentation and communications skills
- Document practice-based learning and improvement in patient care

Journal Club is widely viewed as an appropriate session for teaching concepts in evidence-based medicine.^{1,2,3} We sought to redesign our Journal Club to provide an effective means for assessing our residents' abilities in applying EBM concepts. We first conducted a comprehensive literature review with the goal of establishing "best practices" for the use of Journal Club with outcome measures, and identifying and modifying tools appropriate to the needs of our residency program. We then piloted the revised Journal Club format and tools within the University of Iowa ophthalmology program to determine feasibility, followed by a multi-center trial initiating a process to collect evidence for reliability and validity.

Of the 56 articles we reviewed, 16 reported results of interventional studies on the use of journal club as a teaching tool, using specific outcome measures to evaluate resident performance.⁴ Study designs included cohort, cross-sectional, and pre-and post studies, prospective and randomized trials, and surveys. Outcomes included self and faculty ratings of evidence-based medicine skills using checklists and surveys, knowledge tests, publication of 'letters to the editor', and satisfaction surveys. All but two studies reported improvements in outcome measures.

We used the studies reporting positive findings to develop consensus on "best practices" for using Journal Club as a tool to both teach and assess PBLI in ophthalmology residency programs, listed in **Table 1**. Important changes we made in our Journal Club included:

1. explicit written goals for journal club, communicated to residents and faculty;
2. a formalized meeting structure and process to maximize participation and perceived importance;
3. use of a structured review checklist to assess resident performance outcomes.

The goal in using a standardized checklist was to improve resident performance and consistency of review, provide a template to guide greater attention to detail when reading, and increase perceived satisfaction and self-assessed competency in EBM.

Table 3

Standardized Checklist of Review Criteria⁹

1. **What type of study is this article?** (consult the definitions in *glossary of Study Design* at <http://www.ajo.com>)
 - a. Randomized or non-randomized clinical trial
 - b. Interventional case series or case report
 - c. Cohort study or case-control study
 - d. Cross-sectional study
 - e. Observational case series or case report
 - f. Experimental study
 - g. Meta-analysis of literature
2. **Review the manuscript sections**
 - a. Title: Is the title accurate, concise, and complete?
 - b. Introduction: Are the purposes of the study, the research rationale, and the hypothesis described? Is the pertinent literature reviewed and cited accurately?
3. **Design**
 - a. Methods: Is the description of the study methodology accurate, complete, and appropriate? Does the methods section inadvertently contain results or discussion? Do the methods adequately describe the 1) *Setting* (multi-center, institutional, referral, academic, or clinical practice); 2) *Patients or Study Population* including patient numbers, one or both eyes, selection procedures, inclusion/exclusion criteria, randomization, allocation and masking; 3) *Intervention or Observation Procedure(s)*: (treatments and controls); and 4) *Main Outcome Measure(s)*: (primary, secondary, other).
 - b. Human Subject Participation in Experimental Investigations: Does the manuscript describe the approval from the appropriate Institutional Review Board (IRB) or equivalent monitoring agency? Was appropriate informed consent obtained from the patients or subjects? Does the research conform to generally accepted scientific principles embodied in the World Medical Association Declaration of Helsinki (revised 1989)?
 - c. Use of Animals in Biomedical Research: Does the manuscript describe the animal care protocol, name the institution that sponsored the study, and identify relevant IRB approval? Does the research conform to generally accepted principles of animal maintenance and care (Association for Research in Vision and Ophthalmology guidelines)?
 - d. Statistics: Was the statistical analysis valid? When P values are used, is the actual P value (for example $P = .032$) provided or is an inequality used (for example $(P < .05)$)? In the reporting of the basic summary statistics, are the mean and standard error, as well as confidence limits, provided to help the reader understand the conclusions of the study? Are the statistical models used (analysis of variance, covariance, multiple regressions) specified?
 - e. Results: Are the outcomes and measurements provided in an objective sequence? Are the data provided in a clear and concise manner? Do the tables and figures accurately summarize the data or add to the information presented in the text? Does the data report confidence intervals (usually at the 95% interval) and exact P values or other indications of statistical significance?
 - f. Discussion: Does the discussion accurately describe the results? Does it identify any statistically or clinically significant limitations or qualifications of the study? Do the authors accurately state the conclusions of the study? Are there overgeneralizations or undue speculations in the discussion? Is equal emphasis given to positive and negative findings?

The goals we identified for our structured Journal Club (Table 2) were used in combination with a checklist of review criteria that we developed (Table 3). The criteria were based on those published by the Task Force of Academic Medicine and the GEA-RIME Committee⁵ and adopted by the editorial board of the American Journal of Ophthalmology for review and evaluation of submitted manuscripts.⁶ The local pilot test of the structured Journal Club involved twelve ophthalmology residents at the University of Iowa program. In an informal survey, residents indicated that the process was helpful in establishing a standard format for review of journal club presentations and appreciation of a meaningful mechanism for documenting their participation (written reviews as a portfolio entry rather than simply recording attendance).⁷

We implemented this structured Journal Club over a one-year period at three academic tertiary care ophthalmology programs (The University of Iowa Hospitals and Clinics, The University of Cincinnati, and The University of California, Los Angeles) and evaluated its impact using a pre-post resident self-assessment of domains of PBL (Table 4). Twenty-nine residents in all three years of training (PGY 2, 3, and 4) participated. We found a statistically significant improvement in resident self-assessment scores in PBL in all five PBL domains.⁸

Although we found that the restructured Journal Club was a useful tool for teaching and assessing competence in PBL, our ultimate goal is to demonstrate actual practice-based improvement in resident performance of evidence-based

Table 4
Pre and Post Self-assessment of PBLI Outcomes⁴

Scoring rubric

1. I was unaware of this particular concept prior to the journal club tool being implemented on practice based learning
2. I was aware of this concept but did not apply in the old journal club method
3. I applied the concept during the old journal club but did not document the results
4. I documented application of the concept during the old journal club but did not sustain over time
5. I documented application of concept over time in patient care with improved patient outcomes based upon the relevant practice based learning concept

Five domains of PBLI

- A. To appraise and assimilate scientific evidence from journal articles at journal club in order to improve practice based learning and lead to practice based improvement of patient care over time
- B. To read critically a journal article and draw conclusions applicable to clinical care
- C. To use a systematic and standardized checklist to analyze the paper
- D. To apply knowledge of study designs and statistical methods to the appraisal of clinical studies
- E. To maintain a self-documented written record (e.g., learner portfolio) to use for future improvement of patient care based on the relevant literature examined during journal club

medicine in the “real world”. We plan on expanding our work on journal club to integrate EBM into our curriculum plan. Specifically we will: 1) require residents to assess the impact of a journal club paper on the actual clinical care of real patients over time, using self-selected chart review and documentation in the resident portfolio; 2) require that journal club references and search methodology be included in all resident PowerPoint® presentations at grand rounds; 3) conduct an external review of resident charts (e.g., chart audit) to link changes in EBM behavior over time with recommendations from the journal club; and 4) require that literature reviews connected with specific patient case-logs be included in resident portfolios to document change in practice of EBM over the residency. ■

The University of Iowa Department of Ophthalmology Task Force on the Competencies members include: H. Culver Boldt, MD, Associate Professor; Thomas A. Oetting, MD, Associate Professor; Hilary A. Beaver, MD, Assistant Professor; Richard J. Olson, MD, Assistant Professor; and Keith Carter, MD, Professor and Chair, University of Iowa Department of Ophthalmology.

Also contributing to this work were: M. Bridget Zimmerman, PhD, Biostatistician, University of Iowa; Karl C. Golnik, MD, Program Director and Professor of Ophthalmology, The University of Cincinnati, Cincinnati, Ohio; Anthony C. Arnold, MD, Program Director and Professor of Ophthalmology, The Jules Stein Eye Institute, UCLA, Los Angeles, CA; and Mark C. Wilson MD., Director of Graduate Medical Education, The University of Iowa Hospitals and Clinics, Iowa City, Iowa.

This work was supported in part by an unrestricted grant from Research to Prevent Blindness, Inc., New York, NY.

- 1 Kersten HB, Randis TM, and Giardino AP. Evidence-based medicine in pediatric residency programs: where are we now? *Ambul. Pediatr.* 2005; 5(5): 302-5.
- 2 Kuhn CJ, Wyer PC, Cordell WH, et al. A survey to determine the prevalence and characteristics of training in evidence-based medicine in emergency medicine residency programs. *J. Emer. Med.* 2005; 28(3): 353-9.
- 3 Milbrandt EB and Vincent JL. Evidence-based medicine journal club. *Critical Care.* 2004; 8: 401-2.
- 4 Lee AG, Boldt HC, Golnik KC, et al. Using the journal club to teach and assess practice based learning and improvement in ophthalmology residency training: A review and recommendation for implementation. *Surv. Ophthalmol.* 2005; 50: 542-8.
- 5 Task Force of Academic Medicine and the GEAR-RIME Committee. Checklist of review criteria. *Acad. Med.* 2001; 76(9): 957-959.
- 6 Editorial Manager for Online Submission, American Journal of Ophthalmology. Checklist of Review Criteria: <http://www.editorialmanager.com/ajo/> (accessed 1/4/06)
- 7 Lee AG. Using Journal club as an approach to teaching and assessing practice based learning and improvement. <http://www.acgme.org/outcome/implementation/rsvp.asp> (accessed 1/4/06).
- 8 Lee AG, Golnik KC, Arnold T, et al. Use of the structured journal club to teach and assess practice based learning. *Ophthalmology* (in press).
- 9 Lee AG. Using the American Journal of Ophthalmology's website for assessing residency subcompetencies in practice-based learning. *Am. J. Ophthalmol.* 2004; 137: 206-7.

Morbidity and Mortality Conference: A Practice-Based Learning Tool for the Performance Improvement of Residents and Residency Programs

Joel C. Rosenfeld MD, MEd, FACS

The ACGME Outcome Project established six general competencies (patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice) by which residency program directors are to evaluate the performance of their residents. Program directors are expected to develop assessment instruments to determine the attainment of competency-based outcomes by their residents. Beginning July, 2006, the accreditation focus will include review of evidence that programs are making data-driven improvements for both individual resident and overall residency program performance. Within the St. Luke's Hospital general surgery residency program, we modified our traditional Morbidity and Mortality (M&M) conference with the goal of improving resident performance and then examined the effects of this change on both resident and residency program performance.

“The M&M conference, if properly structured and conducted, can meet all the sub-competencies of practice-based learning and improvement as well as many of the objectives of the other five general competencies and therefore can be utilized both to teach and to assess the competencies.”

Why focus on M&M?

Although the six general competencies are often considered as individual entities by program directors for purposes of implementation and assessment, they are related to each other through the central competency of practice-based learning and improvement. The M&M conference, if properly structured and conducted, can meet all the sub-competencies of practice-based learning and improvement as well as many of the objectives of the other five general competencies and therefore can be utilized both to teach and to assess the competencies.

The M&M conference traditionally is centered on the ACGME general competencies of patient care and medical knowledge. The conference has been referred to as the “golden

hour of surgical education” and the “premier educational conference in most surgery departments”.^{1,2} It has also been referred to as a significant cultural ritual in the development of a surgeon.^{3,4} The conference should enable physicians, both faculty and residents, to discuss adverse events so they can improve their performance and hopefully prevent similar problems from occurring in future patients, thus achieving learning outcomes for practice-based learning and improvement. Research studies have questioned whether the M&M conference really does result in performance

“Unless a resident purposefully reflects about the complication or death he or she presents, they may not really learn from the case.”

improvement.^{5,6} Unless a resident purposefully reflects about the complication or death he or she presents, they may not really learn from the case. Also, if the atmosphere of the conference is one of “blame and shame”, errors and adverse events may not be discussed in an open educational format conducive to learning, one of the key functions proposed for the M&M conference.⁷

Structure of the modified M&M conference

We modified the format of our weekly surgery M&M conference and developed a post conference resident exercise to enable us to use the conference to teach and assess residents in terms of various sub-competencies from all the ACGME general competencies. The focus of our M&M conference is the education of residents, faculty attending surgeons, and medical students. We have a separate performance improvement committee. The basic changes in our M&M format include the following:

1. Each Monday, the chief resident on each teaching service submits a list of patient discharges and deaths from the previous week. Complications, if any, are listed for each individual patient. All deaths are automatically selected for review at the following week's M&M conference as well as all significant patient complications.
2. The resident, who is primarily responsible for the care of the patient selected for presentation, prepares the case for the next week's M&M conference. Besides discussing the case with the responsible faculty attending surgeon, the resident is expected to check hospital specific and/or national databases and treatment protocols that are applicable to the type of patient as well as conduct a literature review of the patient's medical problem and complication. The resident then prepares a PowerPoint® presentation summarizing the patient's hospital course with pertinent laboratory studies and x-rays.

Table 1

Program Improvements Due to Changes in M&M Conference

- Ongoing identification of curriculum weaknesses or deficiencies
- Added new lectures, symposia, interactive exercises dealing with end-of-life issues, medical futility, effective communication skills
- Added sessions on multiculturalism and diversity issues presented by faculty from a local university familiar with the Latino community
- Established an in-hospital course on medical Spanish given by faculty from a local university
- Resident-conducted projects on systems problems identified through M&M that have had an impact on the institution:
 - fail safe system for timely and appropriate review of patient in-hospital laboratory, radiology and pathology results;
 - anticoagulation pathway for clinic outpatients;
 - protocol for use of bedside ultrasound for placement of central venous catheters to decrease the incidence of pneumothorax in difficult patients
 - protocol for aspiration prevention in post-operative patients
 - protocol to decrease the rate of normal appendectomies
- Evidence of scholarly activity by residents (presentations/publications resulting from projects)
- Increased faculty attendance and participation in M&M conference

3. When presenting the case the resident not only discusses the care provided, but also any health care systems problems that contributed to the patient's morbidity and or mortality, patient safety concerns, communications problems involving the patient or health care workers, and any ethical or ethno-cultural issues that may have affected the patient care. Information from appropriate hospital and national data bases is presented as well as an evidence based literature review.
4. The case is then opened for discussion by a previously selected moderator.

“The goal of our M&M conference is education and consequently the conference is conducted in a non-intimidating, non-judgmental, and non-punitive manner.”

At the conference, not only is the patient's care critiqued, but the group also analyzes systems problems, communication problems, possible ethical dilemmas or ethno-cultural issues that may have affected the patient's care. By assessing hospital and applicable national databases and protocols, the resident can compare his/her patient's outcome and treatment result to a larger population of similar patients' benchmarked results. The goal of our M&M conference is education and consequently the conference is conducted in a non-intimidating, non-judgmental, and non-punitive manner.

Post-conference Practice-based Improvement Log

Each resident who presents a case at the M&M conference completes a Practice-based Improvement Log. This form, which we adapted and modified from a form developed by the Department of Surgery at Southern Illinois University School of Medicine, analyzes the patient's hospital course and outcome.⁸ After completing the section describing the possible causes of the complication(s), and/or death, the resident discusses 1) factors contributing to the complication(s), 2) opportunities for systems improvement and for enhancing patient safety, 3) ethical and ethno-cultural issues involved in the care of the patient, and 4) what the resident would do differently in the future as a result of this experience. The resident also lists evidence based references which he/she consulted for the case. The completed form⁹ becomes part of the resident's portfolio and is reviewed with the resident by the residency director.

Program improvements resulting from the modified M&M conference

As a result of the changed educational milieu of our M&M conference, there has been an increased discussion of adverse events and potential problems. Since all factors affecting the patient's care are considered, there is an increased awareness by residents and faculty of the effect of systems problems, communication problems, ethical dilemmas, and ethno-cultural differences on patient care. Weaknesses in our residency curriculum, not only in terms of the competencies of medical knowledge and patient care, but also in the competencies of professionalism, interpersonal and communication skills, and systems-based practice, have been unmasked.

Major program improvements resulting from the modified M&M conference are listed in **Table 1**. As a result of the deficiencies uncovered at the conference, we have instituted lectures, symposia, and interactive resident exercises dealing

“Because of occasional communication problems with members of our large Latino community, we have had faculty from a local university present sessions on multiculturalism and diversity and how different ethnic groups approach their health care. An in-hospital course in medical Spanish for residents and other health care professionals has been established.”

with end of life issues, medical futility, and effective communication skills (written and oral) for communicating with patients and their families as well as medical colleagues. Because of occasional communication problems with members of our large Latino community, we have had faculty from a local university present sessions on multiculturalism and diversity and how different ethnic groups approach their health care. An in-hospital course in medical Spanish for residents and other health care professionals has been established. It is conducted by faculty from a local university.

Systems problems discussed at the conference have resulted in projects by residents and faculty attending surgeons to correct these problems and enhance patient safety. Based on their experiences with patients and what they have considered to be systems problems in our institution, residents working with faculty have developed or are in the process of developing: a fail safe system for patient in-hospital laboratory, radiology, and pathology results to get to the appropriate clinic resident physician for review so that significant abnormal

“Stimulated by discussions at the M&M conference, residents and faculty have also instituted protocols for aspiration prevention in post-operative patients and a pathway to decrease the rate of normal appendectomies. In addition to improving patient care, these projects have resulted in presentations and/or publications by the involved residents.”

results are not missed when patients are discharged from the hospital to the clinic; an anticoagulation pathway for clinic out-patients; a protocol for the use of bedside ultrasound for the placement of central venous catheters in difficult patients to decrease pneumothoraces. Stimulated by discussions at the M&M conference, residents and faculty have also instituted protocols for aspiration prevention in post-operative patients and a pathway to decrease the rate of normal

appendectomies. In addition to improving patient care, these projects have resulted in presentations and/or publications by the involved residents.

Also as a result of the emphasis on education at our M&M conference, there has been increased attendance by faculty attending surgeons and other physicians. The evaluation forms completed by conference attendees demonstrate greater satisfaction with the conference since the introduction of the new format.

Conclusion

By restructuring our M&M conference to allow all competencies to be addressed in an educationally friendly manner, and by implementing a post-conference practice-based learning and improvement tool for residents, we use the weekly M&M conference to improve the performance of our residents. The residency curriculum has been strengthened and some hospital systems problems have been addressed. ■

Joel Rosenfeld MD, MEd is Director of Medical Education and General Surgery Residency Program Director at St. Luke's Hospital & Health Network, Bethlehem, PA. For a copy of the Practice-based Improvement Log, please see reference 9 or e-mail the author at rosenfj@slhn.org.

- 1 Gordon LA. Gordon's Guide to the Surgical Morbidity and Mortality Conference. Philadelphia, PA: Hanley & Belfus, Inc; 1994.
- 2 Sachdeva AK, Blair PG. Educating Surgery Residents in Patient Safety. Surg. Clin. N. Am. 2004; 84:1669-1698.
- 3 Gawande A. Complications: A Surgeon's Notes on an Imperfect Science. New York, NY: Henry Holt and Company; 2002.
- 4 Bosk CL. Forgive and Remember: Managing Medical Failure. Chicago, IL: Chicago University Press, 1979.
- 5 Harbison S, Regehr G. Faculty and Resident Opinions Regarding the Role of Morbidity and Mortality Conference. Am. J. Surg. 1999; 177:136-139.
- 6 Risucci DA, Sullivan T, DiRusso S, Savino JA. Assessing Educational Validity of the Morbidity and Mortality Conference: A Pilot Study. Curr. Surg. 2003; 60(2): 204-209.
- 7 Russell J. Patient Safety and GME Curricula: Function Follows Forms. ACGME Bulletin: August 2004; 11-12.
- 8 Williams RG, Dunnington GL. Accreditation Council for Graduate Medical Education Core Competencies Initiative: The Road to Implementation in the Surgical Specialties. Surg. Clin. N. Am. 2004; 84:1621-1646.
- 9 Rosenfeld JC. Using the Morbidity and Mortality Conference to Teach and Assess the ACGME General Competencies. Curr. Surg. 2005; 62(6): 664-669.

Toward a Broader Understanding of Interpersonal and Communication Skills in Resident Education

Barbara Joyce, PhD

The terms ‘interpersonal’ and ‘communication’ skills are often used interchangeably but the skill sets associated with each of these terms is quite different. Duffy et al.¹ provided a definition that outlines a distinction between communication skills and interpersonal skills: *“Communication skills are the performance of specific tasks and behaviors such as obtaining a medical history, explaining a diagnosis and prognosis, giving therapeutic instructions, and counseling. Interpersonal skills are inherently relation and process oriented; they are the effect communication has on another person such as relieving anxiety or establishing a trusting relationship.”* Curriculum and assessment tools in the area of **Communication Skills** might involve teaching and evaluating residents on such skill sets as conducting the medical interview, sharing bad news, shared decision-making, and informed consent. On the other hand, curriculum and assessment tools in the area of **Interpersonal Skills** might involve a longitudinal assessment of the resident’s ability to sustain a relationship with a patient, their ability to balance medical knowledge with relationship factors involved in patient care, or in forming and sustaining effective team function. In this article, I describe the development of the foundation for all the current efforts to teach and assess communication skills and highlight some issues related to resident education in interpersonal as well as communication skills that program directors and faculty might explore.

The following four papers provide a rich foundation in helping to define the Communication Skills competency and for determining curriculum and assessment tools in this area:

- Kalamazoo Consensus Statement, which defined the essential elements of doctor-patient communication²
- Kalamazoo II, which outlined assessment tools and methods for evaluating communication skills¹
- The Macy Model, which outlined recommendations for teaching communication skills³
- The AAMC Project on the Clinical Education of Medical Students, which outlined communication competencies for medical students.⁴

The Kalamazoo Consensus Statement, reflecting a broadly representative group of experts, educators, leaders, and stakeholders at all levels of medical education, grew out of a synthesis of five common communications models (Bayer Institute for Health Care Communication E4 Model; Three Function Model/Brown Interview Checklist; Calgary-Cambridge Observation Guide; Patient Centered Clinical Method; Segue Framework for teaching and assessing

communication skills).² It identifies the seven essential elements for effective doctor-patient communication, particularly in the primary care ambulatory setting. (Table 1) The seven essential elements have been used to develop, implement, and evaluate communication-oriented curriculum and to inform definition of skills sets within the domain of communication in many medical school settings. Implicit in this model is the assumption that as learners progress through medical school and residency, their competency in communication skills increases from being able to manage generic communication tasks to communication skills that reflect complex specialty specific scenarios.

In 2004, Kalamazoo II took place, involving a different but overlapping group of experts, educators, leaders, and stakeholders with the goal to identify methods for the assessment of communication and interpersonal skills.¹ Three types of common assessments tools were recommended:

1. Focused (direct) observation
2. Patient questionnaire and surveys
3. Examination of knowledge, skills and attitudes

Direct observation tools provide faculty with an opportunity to directly observe and comment on a resident’s skill in communication and might entail use of such tools as the Mini-CEX or Kalamazoo Communication Checklist. Immediate formative feedback about communication skills can be given

Table 1
Seven Essential Elements of the Doctor-Patient Relationship²

- Build the doctor patient relationship
- Open the discussion
- Gather information
- Understand the patients' perspective
- Share information
- Reach agreement on problems, plans and procedures
- Provide closure

to the resident and discussion between resident and faculty can help identify key learning points and activities for improvement. The use of patient questionnaires and surveys, such as the ABIM Patient Satisfaction questionnaire or other patient satisfaction inventories, provide the resident with information on how patients perceive the resident’s communication skills. The ABIM Clinical Skills Module⁵ presents video enactments of a variety of physician-patient encounters along with multiple-choice questions to test knowledge about various communication principles. This format would be useful for residents as well.

The model developed by the Macy Initiative in Health Care Communication (Table 2) is applicable for all specialties and addresses three domains of communication

skills: communication with the patient; communication about the patient; communication about medicine and science.³ The Macy Initiative involved a collaborative effort between three medical schools (New York University School of Medicine, Case Western Reserve School of Medicine, and

“The most significant teaching and learning in the area of Interpersonal and Communication skills is likely to occur at the bedside, in the surgical suite or in the ambulatory clinic where residents have a chance to observe and model senior faculty interacting with patients and staff.”

University of Massachusetts Medical School) to define a framework for development of curriculum and evaluation tools at the medical school level. **Table 3** demonstrates some examples of how diverse specialties might apply the Macy Model to develop curriculum modules in the three domains of communication skills.

The most significant teaching and learning in the area of Interpersonal and Communication skills is likely to occur at the bedside, in the surgical suite or in the ambulatory clinic where residents have a chance to observe and model senior faculty interacting with patients and staff. Branch et al.⁶ outlined three clinical teaching strategies for teaching/ precepting interpersonal as well as communication skills:

- Taking advantage of **seminal events** that encourage discussion about the patient’s experience and relevant psychosocial aspects;
- **Role modeling** where faculty explicitly describe verbal and nonverbal techniques they might use to convey bad news or negotiate a treatment agenda;
- Active Learning methods that encourage **mentor-resident reflection** in the area of psychosocial aspects of patient care.

Encouraging an active dialogue between preceptors and residents around psychosocial issues relevant to patient care is important to facilitating residents’ awareness of how the patient perceives the clinical encounter. Preceptors should be encouraged to go beyond asking residents for the relevant differential diagnosis and treatment plan and begin to ask questions that help deepen the resident’s understanding of the whole person. Examples of such questions might be the impact of cultural issues on patients’ adherence to treatment plans, the role of depression in the patient’s current diagnostic picture, patients’ concerns and insecurities that may influence outcome, or family and other social support. Encouraging residents to engage in self-reflection around biopsychosocial issues related to patient care also deepens their understanding of the patient as a person.

The AAMC Task Force on the Clinical Skills Education of Medical Students recommended two levels of communication skill competence, core and advanced.⁴ Core competencies should be introduced at the beginning of the curriculum and medical students required to demonstrate competence in these skills prior to clerkships. Examples of core competencies include: building the doctor patient relationship, opening the discussion, gathering information and understanding the patient perspective. As medical students progress in their education, they should be expected to demonstrate more advanced competency in such areas as communicating bad news, obtaining informed consent, and counseling for behavior change. A more detail listing of the communication competencies recommended for medical students is contained within the task force report and may be useful for program directors and faculty to review.

Wide variability currently exists in medical students’ experiences in communication training. Medical educators might assume that residents have learned communication skills in medical school, based on the earlier work and recommendations just described, however, it is important for program directors and faculty to explore whether their residents actually can demonstrate those foundational communications skills. Duffy et al.¹ described how many

Table 2

Macy Model of Doctor Patient Communication*

Communication with the Patient

Communication in the doctor patient relationship

- Conducting the medical interview
- Discussing treatment options

Content specific areas

(could reflect specialty specific communication modules):

- Breaking bad news
- Organ/tissue donation
- Informed consent
- Family interview
- Shared decision making

Communication about the Patient

Oral communication

- Case presentations
- M & M
- Consultation
- Hand-off

Written communication:

- H & P
- Pre/post op note
- Consultation note
- Transfer and discharge

Team Communication

- Multidisciplinary treatment planning teams
- Coordinating care with ancillary staff or different units
- In the operating suites or labs; transfer of care; hand-offs
- Leadership; conflict resolution

*Adapted from Kalet, et al.³

Table 3

Application of the Macy Model

	Medical Specialties	Surgical Specialties	Hospital Based Specialties	Assessment
Communication with the Patient	Medical Interviewing	Informed consent	Medical interviewing	Direct observation checklists
	Breaking Bad News	Breaking bad news		Patient satisfaction surveys
	Shared decision making re: treatment options	Organ tissue donation		Examination of knowledge, skills and attitudes
Communication about the Patient	Holding Effective family meetings	Team building Hand-off	Team building Leadership	Multi-source Professional Assessment (360)
	Hand offs	Effective family meetings		
	Consultation-liaison	Case presentations		
	Case presentations	M & M		
Communication about Medicine	Grand Rounds	Grand Rounds	Grand Rounds	Evaluation of research project and/or professional presentations
	Scholarly presentations	Scholarly presentations	Scholarly presentations	

programs use observed encounters with standardized patients, focused on physical examination and communication skills, during the first year of training as a formative evaluation of an entering resident's communication skills. In some programs, residents received detailed feedback from faculty as well as an audiotape of their performance. Repeating the same exercise at the end of residency can serve as a summative evaluation.

When viewed from these perspectives, Interpersonal and Communication Skills are more than medical interviewing and history taking. These skills, along with medical knowledge and patient care skills, form the bedrock of patient care, contribute to physician and patient satisfaction, and contribute substantially to treatment adherence.

Because developing expertise in these skills is an ongoing, lifelong process, based fundamentally in experience, program directors need to determine where in the continuum each resident is on entering the program. Teaching interpersonal and communication skills needs to go beyond a yearly workshop or a noon lecture and be incorporated into the daily work done with residents, with specific attention to encouraging mentor-resident reflection in the psychosocial aspects of patient care. ■

Barbara Joyce, PhD, is a Senior Project Manager in the Research and Education Dept. at the ACGME, Chicago, IL. Prior to joining the ACGME, Dr. Joyce was an Educational Consultant to the Henry Ford Health System in Detroit MI, where she was responsible for implementing the Outcome Project.

- Duffy FD, Gordon G, Whelan G, Cole-Kelly K, Frankel R. Assessing Competence in Communication and Interpersonal Skills: The Kalamazoo II Report. *Acad. Med.* 2004; 79(6): 495-507.
- Participants in the Bayer-Fetzer Conference on Physician-Patient Communication in Medical Education. Essential Elements of Communication in Medical Encounters: The Kalamazoo Consensus Statement. *Acad. Med.* 2001; 76(4): 390-393.
- Kalet A, Pugnaire M, Cole-Kelly K, Janick R, Ferrara E, Schwartz M, Lipkin M, Lazare A. Teaching Communication in Clinical Clerkships: Models from the Macy Initiative in Health Communications? *Acad. Med.* 2004; 79: 511-520.
- AAMC Task Force on the Clinical Skills Education of Medical Students: AAMC Project on the Clinical Education of Medical Students: Clinical Skills Education. AAMC, Washington, DC. 2005.
- American Board of Internal Medicine Self-Evaluation of Medical Knowledge: Clinical Skills Module: <http://www.abim.org/moc/semmed.shtm#5> (Accessed 2/24/06).
- Branch W, Kern D, Haidet P, Weissmann P, Gracey C, Mitchell G, Inui T. Teaching the Human Dimension of Care in Clinical Settings. *JAMA* 2001; 9: 1067-1074.

The Effect of Standardized Patient Feedback in Teaching Surgical Residents Informed Consent

Kendall Reed, DO, FACOS

Informed consent is a complex process that represents an important interaction between a patient and physician and, as such, is a critical **competency** for all physicians regardless of specialty. The informed consent process usually consists of a verbal discussion, which is frequently accompanied by the patient's signature on a written document that verifies the patient has been informed of options, possible complications and outcomes regarding certain diagnostic studies, procedures or treatment modalities. This interaction, ideally, should satisfy the physician's moral and legal responsibility while simultaneously increasing the patient's satisfaction with, respect for, and trust in the physician. If informed consent is done correctly, the desired result should be decreased malpractice liability and enhanced physician-patient relationships.

Because of the paucity of available literature on informed consent in the surgical literature coupled with the need to educate our surgical residents on this most important competency, we wanted to determine the effectiveness of using **feedback** from standardized patients (SP) who were trained to simulate four common surgical problems of increasing complexity (appendectomy, cholecystectomy, colon resection and surgical options in the treatment of breast cancer).¹ Eight volunteer residents in PGY 2–5 years (two at each PGY level) participated with four assigned to the group receiving feedback and four to a group that did not receive feedback. All of the residents were required to do two SP encounters with each of the four case types. A total of 16 SP's were used in this project. All resident/SP encounters were videotaped and served as a self-assessment tool for each resident between encounters.

The "feedback" group received a feedback debriefing by the SP between the first and second encounter of each major case type. In addition, this group viewed their videotape with the SP and then completed a self-assessment prior to receiving the SP feedback. The resident self-assessment form was identical to the SP feedback form, and included 14 questions focusing on establishment of patient rapport, communication skills, technical considerations/complications, and closing the session.¹ The "no feedback" group completed all of the above but did not receive feedback from the SP. Different SP's were used in the two encounters for each case type and the SP's were blinded as to which resident had received feedback after the initial encounter. The data retrieved from the 14-item assessment instrument used by the SP's to evaluate the residents on the informed consent process was analyzed using ANOVA, including F-tests

(univariate and multivariate) for 1) group effect, 2) time effect (pretest to posttest and over the various cases), and 3) group X time interaction. The results revealed a statistically significant overall change- pretest to posttest and across cases for the "feedback" group ($p < 0.01$). The group effect was also statistically significant ($p < 0.01$) with the "feedback" group averaging 10 points greater than the "no feedback" group.

The use of videotaped encounters with standardized patients is becoming a more widely used tool for teaching and assessing essential outcomes in the competency domain of interpersonal and communication skills. Burchard, et al described its use as an assessment method for interpersonal

"We believe that standardized patients and their feedback is a potentially effective modality to assist in teaching residents the process of informed consent, although the conclusions in our study are limited by small sample size."

skills of third year medical students.² The development of an instrument to measure doctor-patient nonverbal communication using videotaped SP encounters with medical students was described by Gallagher, et al.³ Rosengren, et al developed a video assessment of simulated encounters (VASE) as a method for evaluating clinician skills in motivational interviewing.⁴ We believe that standardized patients and their feedback is a potentially effective modality to assist in teaching residents the process of informed consent, although the conclusions in our study are limited by small sample size. This is a critical competency which we continue to refine and explore for both residents and medical students. ■

Kendall Reed, DO, is Dean and Professor of Surgery at the College of Osteopathic Medicine at Des Moines University, Des Moines, IA.

¹ Leeper-Majors K, Veale JR, Westbrook TS, Reed K. The effect of standardized patient feedback in teaching surgical residents informed consent: results of a pilot study. *Current Surgery* 2003; 60(6): 615-622.

² Burchard KW, Rowland PA, Berman NB, Hanissian PD, Carney PA. Clerkship enhancement of interpersonal skills. *Amer. J. Surg.* 2005; 189: 643-646.

³ Gallagher TJ, Hartung PJ, Gerzina H, Gregory Jr SW, Merolla D. Further analysis of a doctor-patient nonverbal communication instrument. *Patient Education and Counseling*. 2005; 57: 262-271.

⁴ Rosengren DB, Baer JS, Hartzler B, Dunn CW, Wells EA. The video assessment of simulated encounters (VASE): Development and validation of a group-administered method for evaluating clinician skills in motivational interviewing. *Drug and Alcohol Dependence*. 2005; 79: 321-330.

Asking to Learn: A Study of Performance Outcomes for Program Improvement

George R. Bergus, MD, MA and Myra Emerson, MA, JD

Learning through problem solving is an activity that routinely occurs during the workday of any practicing physician. Clinical questions, which represent knowledge gaps, frequently arise in the course of delivering clinical care. While physicians may use books, computer programs, and websites to answer these questions, they frequently turn to a more experienced colleague for help. Often this request is an informal “curbside consult.” Primary care physicians have identified this type of interaction to be one of the most common reasons for changing their practices.¹

“We found that when physicians translate their clinical problems into well-formulated questions, expert consultants were more likely to provide information and suggestions that helped the physician manage the problem rather than to simply recommend a formal consult.”

Decades ago, Vygotsky proposed that the lifelong process of development is dependent on social interactions which result in cognitive learning through mentoring from individuals who are more expert.² Learning occurs within ‘zones of proximal development’ – gaps between what can be done with help and what can be done without guidance. Physicians experience zones of proximal development when they seek expert help to solve clinical problems by requesting informal consultations. These interactions not only allow patients to receive good medical care but they can also result in physician learning. We believe the ability to enlist expert help is important for physicians attempting to capture the learning opportunities in their daily clinical practices.

We reviewed five years of clinical question and answer exchanges between primary care and specialty physicians in order to determine how effective primary care physicians – including family medicine residents – were in learning from their daily clinical practices. We found that when physicians translate their clinical problems into well-formulated questions, expert consultants were more likely to provide information and suggestions that helped the physician manage the problem rather than to simply recommend a formal consult. On the whole, residents in our family medicine program formulated good, but not excellent, questions. However, individual residents did not improve this skill during training. These results suggest an area for program improvement.

In 1996 we created an electronic mail-based informal consultation service connecting selected primary care clinicians in Iowa with specialty physicians at the University of Iowa Carver College of Medicine. Between 1996 and 2001, over 80 family physicians used the E-mail Consult Service (ECS) to ask over 2,000 questions.³ The physicians were full-time community practitioners, or faculty members, or residents in family medicine programs. Neither the family physicians nor the specialty consultants were instructed on how to formulate their questions or answers except that they were advised not to include unique patient identifiers in the messages in order to protect patient confidentiality. The median turnaround time between sending a question and receiving an answer was about 17 hours, and nearly 65% of the questions were answered within 24 hours.

Although the family physicians were highly satisfied with the ECS, periodically one would complain that a specialist had recommended formal consultation when the family physician thought the problem was manageable with the informal advice of the specialist. Specialists recommended formal consultation in response to about 9% of the questions.

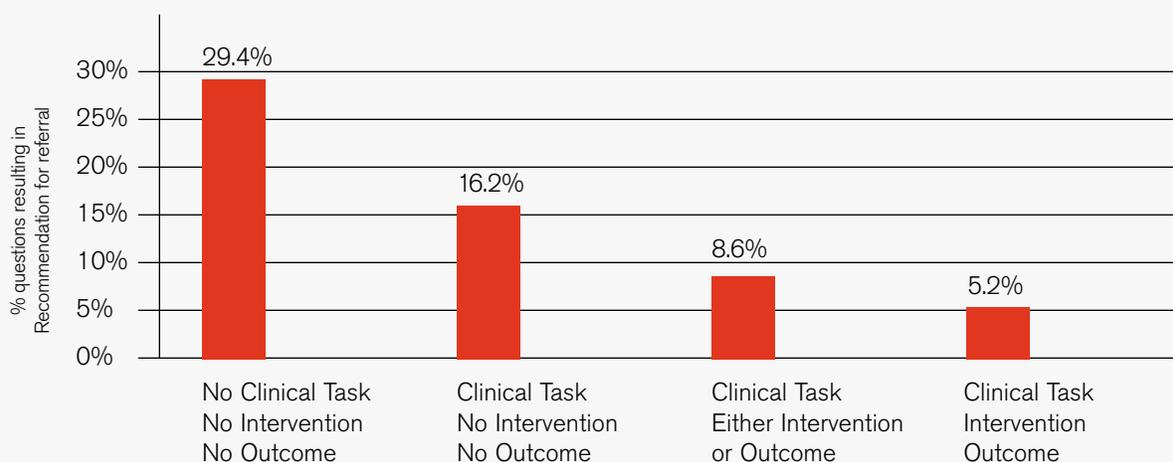
An extensive analysis of these “failed” curbside consults lead us to conclude that how physicians formulated their clinical questions was a major predictor of whether specialists would informally help solve the problem or ask for formal referral.³ We came to this conclusion after analyzing 1,618 clinical questions about specific patients using the PICO taxonomy developed by Sackett and his colleagues.⁴ We parsed each question to identify the four PICO components: the *Patient/Problem*, the *Intervention* of interest, a *Comparison* intervention, and the desired clinical *Outcome*. Clinical questions were also placed into one of six clinical task categories: Diagnosis, Prognosis, Treatment, Prevention, Practice Improvement, and Request for Direction. This last category was used for questions that contained a description of a clinical situation but then did not identify one of the other clinical tasks. Questions in this category were generally in the form of, “What do you think?” or “Any ideas?” or “What would you suggest I do?”

“Using Hierarchical Linear Modeling, we uncovered a very strong relationship between question formulation and the specialist’s recommendation for formal consultation ($p < 0.001$).”

Using Hierarchical Linear Modeling, we uncovered a very strong relationship between question formulation and the specialist’s recommendation for formal consultation ($p < 0.001$). Fewer than 5% of questions in which the primary care physician identified the clinical task of concern, proposed an intervention and identified the desired outcome resulted in a recommendation of formal consultation. In contrast, about 30% of the instances that identified none of these

Figure 1

Clinical Task, Intervention, and Outcome question components and specialists' recommendations for consultation.*



*Differences in rates are statistically significant ($p < 0.0001$). Used with permission.³

question components resulted in this outcome. In addition, there was a monotonic relationship between these three question components and the recommendation for formal consultation (**Figure 1**) allowing us to grade the quality of clinical questions on a 0 to 3 scale. Examples of clinical questions graded for quality are shown in **Figure 2**.

We looked for alternative explanations for the strong association between question quality and the recommendation for formal consultation. Because well-formulated questions had to contain an intervention and outcome, we considered that specialists might simply be confirming a satisfactory plan when one was presented as a well-formulated question. However, we found that about half the time specialists did not endorse the family physician's proposed intervention and this lack of agreement did not seem to influence whether formal consultations were recommended ($p = 0.57$). In addition, the amount of clinical information included in the e-mail curbside consult request did not influence the outcome ($p = 0.22$) after controlling for question quality. Similarly, whether the questioner was a resident or faculty did not appear to influence whether the specialist asked for a formal consultation ($p = 0.61$) after adjustment for question quality. We found that faculty asked significantly better questions than did residents, although the difference was small (quality scores: 2.2 vs. 2.1, $p = 0.02$).

After finding the association between question formulation and a specialist's willingness to informally help a family physician manage an unfamiliar clinical problem, we looked at whether our residents asked better-formulated questions as they progressed in their residency. There are several reasons to anticipate this improvement. Our residents are exposed to many lectures and journal clubs where the principles of evidence-based medicine, including the importance of well-formulated questions, are stressed. In addition, residents spend a great deal of time learning through answering clinical questions about their patients, and it is possible that this ongoing experience results in better skills at question formulation for curbside consult requests.

For this analysis we restricted our data set to the 454 questions asked by residents.⁵ Individual residents asked a median of six questions (range = 1-40). One hundred and ten (24%) of the questions were asked by residents in their first year, 179 (39%) in their second year, and 165 (36%) in their third year. The mean quality of the clinical questions asked by all residents was 2.1. Individual residents had means ranging from 0 to 3, signifying that residents asked questions ranging from unformulated to well-formulated. The mean quality

“Using a repeated measures ANOVA, we did not find evidence that question quality changed as they progressed in their clinical education ($p = 0.86$). We also did not detect an association between question formulation skills and a more general measure of medical knowledge.”

scores of the questions asked by first, second, and third year residents were essentially identical ($p = 0.98$). Additionally, the distribution of the 454 questions by quality scores in each of the three years of residency were also nearly identical ($p = 0.97$). Thus, we did not detect any differences in the quality of the clinical questions formulated by residents over the three years of training when our data was analyzed cross-sectionally.

A subset of the residents (13) used the e-mail service during each of their three residency years, asking a total of 220 questions. These residents showed significant variation in their skills at constructing well-formulated clinical questions ($p < 0.001$). The mean question quality score for these residents was also 2.1, and the individual residents generated mean quality scores ranging from 0.8 to 2.4. Using a repeated measures ANOVA, we did not find evidence that question

Figure 2

Examples of clinical questions categorized according to the PICO taxonomy.*

**No Clinical Task, No Intervention, No Comparison Intervention, and No Outcome question components
(quality score = 0)**

"I have a 67-year-old man who is getting up 3 times a night to void. He is already on trazosin 10 mg a day. BP is 102/70 and without orthostatic symptoms. What do you think?"

**Clinical Task but no other question components
(quality score = 1)**

"I have a 67-year-old man who is getting up 3 times a night to void. He is already on trazosin 10 mg a day. BP is 102/70 and without orthostatic symptoms. What can I do with his medication?"

**Clinical Task and Intervention question components
(quality score = 2)**

"I have a 67-year-old man who is getting up 3 times a night to void. He is on trazosin 10 mg a day. BP is 102/70 and without orthostatic symptoms. Should I add finasteride?"

**Clinical Task, Intervention, and Outcome question components
(quality score = 3)**

"I have a 67-year-old man who is getting up 3 times a night to void. He is on trazosin 10 mg a day. BP is 102/70 and without orthostatic symptoms. Should I add finasteride to improve his obstructive urinary symptoms?"

**Clinical Task, Intervention, Comparison Intervention, and Outcome question components
(quality score = 3)**

"I have a 67-year-old man who is getting up 3 times a night to void. He is on trazosin 10 mg a day. BP is 102/70 and without orthostatic symptoms. Should I add finasteride or increase his trazosin to improve his obstructive urinary symptoms?"

*Used with permission.³

quality changed as they progressed in their clinical education ($p = 0.86$). We also did not detect an association between question formulation skills and a more general measure of medical knowledge. The correlation between the residents' quality of question scores and their in-training exam scores was not statistically significant ($r = 0.08$, $p = 0.80$).

These findings suggest that a physician's success in securing expert help through informal consultation is related to how well the physician formulates the clinical question. In addition, we find a wide range in question formulation skills in our residents but no evidence that this skill improves over the 3 years of residency. Because residents' skill in formulating questions did not improve, many may have graduated without the skills needed to maximize their self-directed learning.

Villanueva, et al found that an intervention targeting literature searches for answers to clinical questions can enhance question formulation skills.⁶ Whether a focused intervention would enhance the question formulation skills of a physician within the venue of informal consultation is, to our knowledge, unstudied. Our residency program includes several lectures about question formulation using the PICO taxonomy. In retrospect it does not seem surprising that we find no evidence that our residents improve their question formulation skills during their education. Passive educational interventions are generally considered ineffective for learning and retaining new skills.^{7,8} Through our study of resident performance and reflection on learning opportunities provided within our program for improving performance in formulating consult questions, we have identified an important area for program improvement. ■

George Bergus, MD, MA, is Professor of Family Medicine in the Carver College of Medicine and Myra Emerson, MA, is a research assistant in the Department of Family Medicine, both at the University of Iowa.

- 1 Allery LA, Owen PA, Robling MR. Why general practitioners and consultants change their clinical practice: A critical incident study. *BMJ*. 1997; 314(7084): 870-874.
- 2 Rieber RW, Carton AS. The collected works of L. S. Vygotsky. Volume I: Problems of general psychology. Plenum Press. New York. 1987
- 3 Bergus GR, Emerson M, Reed DA, Attaluri A. Email teleconsultations: Well formulated clinical referrals reduce the need for clinic consultation. *J Telemed Telecare*. 2006; 12(1): 33-38.
- 4 Sackett DL, Straus S, Richardson S, Rosenberg W, Haynes RB. Evidence-Based Medicine: How to practice and teach EBM (2nd Edition) Churchill Livingstone. London. 2000.
- 5 Bergus GR, Emerson M. Family medicine residents do not ask better-formulated clinical questions as they advance in their training. *Fam Med*. 2005; 37(7): 486-490.
- 6 Villanueva EV, Burrows EA, Fennessy PA, Rajendran M, Anderson JN. Improving question formulation for use in evidence appraisal in a tertiary care setting: A randomised controlled trial. *BMC Med Inform Decis Mak*. 2001; 1: 4.
- 7 Mentkowski, M. Learning That Lasts: Integrating learning, development, and performance in college and beyond. Jossey-Bass. San Francisco. 2000.
- 8 Thomson O'Brien MA, Freemantle N, Oxman AD, Wolf F, Davis DA, Herrin J. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst. Rev*. 2001; (2): CD003030.

The Development and Application of a Competency-based Assessment Tool

Lawrence M. Reich, MD

Defining the need for program improvement

The apprentice model long ago lost its relevance to medical education, amid the expanding volume and complexity of medical knowledge and technology, the growing complexity of medical and social systems, and the evolving societal perceptions of the roles, responsibilities and accountabilities of physicians. From newer models, however, two paradoxes emerge. One, that it is possible for a medical resident to command a great deal of medical knowledge and still not be an effective physician, and two, that it is possible for a resident or fellow to have all the skills necessary to be an effective physician and still not be able to direct these skills towards effective patient care.

The structuring of the domains of medical education into six core competencies in part addresses the first paradox, and the notion that the most appropriate assessment of the outcomes of medical education is the demonstration of actual, not potential, clinical performance addresses the second.

The goals of education now entail, for example, the demonstration that the resident who has mastered the competencies of *medical knowledge* and *patient care* can also overcome the barriers to access (*systems-based practice*), can effectively communicate with the patient (*interpersonal and communication skills*), can attain an effective physician's role by mastering the competency of professionalism, and can apply skills for a lifelong updating of clinical practice (*practice-based learning and improvement*). But demonstration that the resident has mastered these competencies, while necessary, is not a

sufficient goal of medical education. We must ensure the actual effectiveness of training as opposed to its potential effectiveness.

Ensuring the actual effectiveness requires a different set of questions (do residents achieve the learning objectives and how can this be assessed), and a different set of priorities (clearly identifying learning objectives, assessing attainment of these objectives, and using these data to facilitate the continuous improvement of both residents' and the program's performance). With this in mind, we describe the process by which we refocused our priorities at the Mount Sinai School of Medicine (Elmhurst) Program in Internal Medicine, with a view toward the comprehensive and integrative needs and desired outcomes of the residents, the faculty, the program and the institution as a whole.

“Previously, our curriculum had been organized around *content* (what we taught) and *process* (how we taught it). Our goal was to reorganize it around *outcomes*.”

Developing the CEPI Tool

While engaged in the process of revising our written curriculum, we re-examined the role that each programmatic component played in resident education. Previously, our curriculum had been organized around *content* (what we taught) and *process* (how we taught it). Our goal was to reorganize it around *outcomes*. The relevant questions were not “is this what we want the resident to know (*content-based curriculum*)” or “is this the best way to teach it (*process-based curriculum*)”, but rather, “does the presence of this particular content and process in the curriculum help train residents who have mastered the competencies of practice (*outcomes-based curriculum*)”.

We developed an approach we called comprehensive educational performance improvement (CEPI). As a first step, we applied a reductionist perspective, aimed at identifying each distinct learning element of the program, and classifying each element into its clinical, didactic, or evaluative domain. We identified 40 distinct clinical learning elements (specific outpatient and inpatient settings through which the residents rotated or performed service) and 25 didactic learning elements (specific lecture formats, workshops and conferences in which the residents participated). Additionally, we included 11 evaluative elements (evaluation formats and contexts other than those specific to the clinical and didactic learning elements (**Table 1**)). We approached these elements with a similar set of criteria. Are they competency-based, are they the best, most appropriate, or most useful methodologies, and do they further programmatic goals?

Table 1

Evaluative programmatic learning elements

- Nurse Evaluation
- Mini-CEX, Clinical Evaluation Exercise
- Rotation-based Multiple Choice Question Exam
- Evaluation by Ancillary Hospital Staff
- Patient Evaluation
- Peer Evaluation
- Review of Videotaped Patient Encounter
- Evaluation of Research Project
- Chart Audit – Outpatient Progress Note
- Chart Audit – Hospital Discharge Summary
- Monthly Evaluation Form

Table 2

Sample CEPI Page: Clinical Outpatient Service

Component: MPC-Pap Smear Service
 Category: Clinical
 Setting: Outpatient Clinics
 Faculty: Dr. A.
 Description: Residents see MPC patients for cervical cancer screening
 Competencies: All

Question	Assessment	Competency	Personnel	Priority
Do residents find this rotation to be clinically useful?	Resident Survey	All	Chief Resident	
Is there a proper balance between service and learning?	Resident Survey	All	Chief Resident	
At the conclusion of this rotation, does the resident feel competent to perform a Pap smear?	Resident Survey	All	Chief Resident	
Is this an effective modality for learning this content?	Assessment of resident's mastery of this content, assessed by Board performance, in-service exam, or patient outcomes.	All	Dr. A	
Does this rotation have an impact on the resident's professional development?	Reflective learning. "Can you recall a particular experience you had in this rotation that was important to you?"	Professionalism	Dr. A	

Comments:

We then applied to each learning element a set of questions intended to define and to describe, qualitatively and quantitatively, its programmatic value. We addressed the following issues:

- Can this learning element be categorized into one or more of the core competencies? Can its actual effectiveness (its outcome) be described or measured in terms of resident performance? How does this element specifically further the educational goals of the Program?
- Is this the most effective modality for teaching and learning this specific content?
- Does this element provide the proper balance between education and service?
- Does it have a measurable impact on the resident's professional development?
- Is its inclusion in the program consistent with the needs of the institution?

Applying these questions, when appropriate, to each of the 76 distinct programmatic learning elements produced a compendium of 301 questions. The next step, prioritization, was accomplished by asking relevant residents, faculty and staff to assign priority scores for each of them, and to suggest or develop methodologies for addressing them. Using the CEPI process, we created a comprehensive and coordinated plan for the assessment of the program and for the competency-based assessment of the residents. The compendium represented a blueprint of where the program stood, as well as a roadmap of where the program needed to go to achieve a relevant outcomes-based assessment. A sample page from this compendium is shown in **Table 2**.

Application of CEPI for Program Improvement

Case 1. Didactic Small-group Workshop. Residents participate in a monthly small-group Dermatology workshop, a discipline integral to the education of the General Internist. The workshop's format included a review of general topics with discussions of diagnostic and therapeutic management

issues. Some concerns were raised about the value of this workshop in the education of the residents – residents may have difficulty in applying the content in the clinical setting, this may not be the most effective teaching format, and the workshop may not contribute meaningfully to the residents’ overall development. Accepting that merely identifying this clinical

“Some concerns were raised about the value of this workshop in the education of the residents – residents may have difficulty in applying the content in the clinical setting, this may not be the most effective teaching format, and the workshop may not contribute meaningfully to the residents’ overall development.”

content as important knowledge for the resident is insufficient grounds for continued inclusion in the program, the CEPI approach was applied. An instrument to assess objectively the acquisition of medical knowledge in this area (multiple-choice question test) was developed, and external objective assessment (in-service exam performance) was also used.

A questionnaire to assess the residents’ subjective experience in the workshop was developed, asking residents to rank the importance to them of their experiences in the workshop according to the criteria established by the competencies of medical knowledge and patient care. They were also asked to assess the workshop’s value for their professional development and their ability to clinically apply this knowledge. As a result of these assessments, the curricular content was modified and learning objectives congruent with the residents’ needs were established. We had a better sense of the stage of training in which the workshop had the most impact and also of its value in the residents’ professional development. The process created a heightened sense of professional satisfaction for the faculty who were participating in outcomes-based education research and maximizing educational effectiveness.

Case 2. Outpatient Clinical experience. Residents rotate through an outpatient experience in a tertiary referral clinic where they encounter patients who suffer from an uncommon disease. Recognizing that most residents would not, in their clinical careers, be responsible for the management of patients with this condition, concern arose over the programmatic value of this rotation. Applying the CEPI approach, we surveyed residents about their views of the clinical utility of the rotation, their perceptions of the balance between service demands and learning experiences, and the role of the experience in their professional development.

Responses were analyzed as a function of the subjective assessment of each resident’s mastery of the management of patients in this clinical setting (their evaluation), and as a function of an objective assessment of their synthesis of the clinical content (content-based exam). A comprehensive picture of the programmatic value of this rotation emerged, facilitating a refinement and redirection of its learning objectives and curricular goals.

Case 3. Evaluative. We developed an instrument in which ward nurses provided feedback on resident performance in the competencies of *professionalism, systems-based practice, and interpersonal and communication skills*. Recognizing that gathering data was easier than ensuring that the data contributed meaningfully to outcomes-based resident assessment, we invoked CEPI. To answer the question “do the nurses’ evaluations add new information about resident performance?” we compared the nurses’ assessments of these competencies with that of other evaluators (e.g., attending physicians).

We looked for external measures of the attainment of these competencies (i.e., performance on a clinical evaluation exercise and summative evaluation scores) to correlate with the scores given by the nurses. Further, we examined the role of the process of nurse evaluation in improving the quality of interpersonal and professional interactions between residents and nurses. By this approach, a comprehensive view of the overall programmatic value of instituting these evaluations was obtained.

“The Outcome Project provides a framework for assessing and assuring the success of these processes, but how to apply this framework has remained a challenge for program directors.”

Benefits realized through use of the CEPI Tool

Increasing public concern about physician professionalism and accountability has placed scrutiny on the role and quality of medical education. The Outcome Project provides a framework for assessing and assuring the success of these processes, but how to apply this framework has remained a challenge for program directors. Our program analysis resulted in a 76-page compendium of 301 questions that provides us with a handy document that we have used in many ways to improve our program:

- Collect input and feedback from teachers and learners on program goals and directions;
- Create a blueprint of where the program is and a roadmap for where the program needs to go;

-
- Provide a mechanism for the coordinated and comprehensive evaluation and assessment of diverse program elements;
 - Facilitate concurrent assessment of content- and process-based features with outcome-based features of each learning element;
 - Enable simultaneous assessment of a learning element's outcome (demonstration of competency-based effectiveness) and its programmatic value;
 - Integrate assessment of cognitive and clinical aspects of a programmatic element while also integrating the input of evaluators at various levels;
 - Aid in the identification of elements that fill “service demands” but that lack a learning benefit and in the selection of clinical learning environments where specific elements of learning may best occur; and
 - Facilitate systematic evaluation of the progress of learning over the course of the 3-year program.

Additionally, the CEPI tool plays an important and ongoing role in faculty development, namely, orienting broad segments of the faculty towards an evidence-based research mode of thinking and teaching. ■

Lawrence Reich, MD, is Associate Program Director of the Internal Medicine Residency Program at Mount Sinai School of Medicine (Elmhurst), Elmhurst Hospital Center, Elmhurst, NY.

Fostering Self-Assessment and Self-Directed Learning in the Intensive Care Unit

Alison S. Clay, MD

The key to practice-based learning and improvement is accurate self-assessment. Graduate medical education has not traditionally incorporated self-assessment. Though self-assessment is often inaccurate, it is a skill that is not only needed during residency but also essential to the maintenance of knowledge and development of new skills in the practicing physician.^{1,2,3} Accurate self-assessment can be fostered by providing examples for performance, offering opportunities to compare performance to this benchmark and by providing specific feedback about self-assessment.^{4,5} We sought to foster self-assessment by residents in the ICU by developing tools to track and assess resident performance, increasing self-directed learning opportunities and increasing feedback to residents about their performance during their ICU rotation. This project evolved to include a resident portfolio and subsequently the creation of what might be called a “program portfolio.”

We started by developing and validating several assessments specific to critical care medicine, including a 16-item self-confidence assessment, a 45-question medical knowledge test and a 5-item 360-degree evaluation specific to critical care medicine.⁶ We specifically created overlap between the different assessment tools so as to obtain multiple “views” of resident competence in all six competency domains.

“After reviewing published practice guidelines, we also developed 5 checklists that highlighted “expected” care for septic shock, hemorrhagic shock, consultation to the emergency department, discussions with family, and sterile placement of a central line.”

We then developed a learner-directed online course in intensive care medicine. This course consisted of 10 physiology based lectures, each with a multiple choice post-lecture test, a critical care website with useful critical care links and reference list with links to full text articles highlighting evidenced-based medicine in the ICU. After reviewing published practice guidelines, we also developed 5 checklists that highlighted “expected” care for septic shock, hemorrhagic shock, consultation to the emergency department, discussions with family, and sterile placement of a central line. The checklists not only defined our expectations, but also were to be used for “debriefing”.⁷ Following clinical encounters, the fellow and

resident would sit down privately to review how the team performed with respect to practice guidelines, and discuss how they might perform differently in the future. The fellow would give feedback to the resident on the resident's assessment of the team's performance.

At the beginning of the month, all residents took the self-confidence assessment and knowledge test. The residents received feedback based on their results with respect to the core competencies and also with respect to the performance of their peers. At the start of the rotation, half of the residents were randomized to the new curriculum in addition to our

“The residents who had been randomized to the new curriculum also had the opportunity to assess their knowledge and performance during the rotation – using the online lectures and tests to assess and improve their knowledge and using the checklists and debriefing to assess their performance at the bedside.”

standard teaching practices. Use of lectures, post-lecture tests and checklists was tracked during the month. At the end of the month, all residents re-took the self-confidence assessment and knowledge test and were evaluated by the nurses, fellows and faculty using the 360-degree evaluation. Within 10 days of completing their rotation, the residents received the pre- and post rotation scores and their 360-evaluations. This evaluation also included comparisons between medical knowledge, self-confidence and performance as seen by others. The residents who had been randomized to the new curriculum also had the opportunity to assess their knowledge and performance during the rotation – using the online lectures and tests to assess and improve their knowledge and using the checklists and debriefing to assess their performance at the bedside.

At the end of six months, utilization of the online tests and debriefing cards was poor. A full 44% of the residents did not complete even one of the post-lecture self-assessments, and 30% did not complete more than one debriefing card. However, residents who completed more self-assessments had less variance in their evaluations by supervisors, more accurate self-assessment, and were less likely to receive a “less than competent” grade by any evaluator.

Even though many residents did not use the additional learning resources offered to them, simply offering the tools improved resident perceptions of the learning environment. The residents randomized to the new curriculum believed that the ICU environment was more conducive to learning and were less likely to request additional learning opportunities than those residents who were not offered access to these resources.

Furthermore, all the residents were also quite open to feedback. When asked about the utility of the self-confidence assessment, knowledge test, and 360-degree assessment, the residents found all of these resources useful except nurses' evaluations. Anecdotally, the timely receipt of end-of-month evaluations was also appreciated. More than one-third of the residents responded directly to the study coordinator about their performance and 3 residents requested individual meetings to discuss how to improve their performance.

The literature shows that development of expertise requires repeated, deliberate practice with accurate self-assessment.⁵ In our study, the residents who “practiced” through the use of online tests and bedside checklists and who received feedback about these performances became more accurate in self-assessment. They also demonstrated features of expert performers: less variability in performance and fewer unacceptable performances as seen by multiple observers.

We have built on these results by creating an intensive care unit specific website (<http://criticalcare.duhs.duke.edu>) that embraces the multiple disciplines involved in the ICU at our institution (surgery, anesthesiology, and internal medicine). The website includes frequently referenced articles with full text links, the online lectures (and others) that were used in this study, and a resident portfolio that was created specifically for critical care medicine. By utilizing the expertise of several different disciplines, we were able to create the portfolio in a timely fashion, minimize redundancy between programs and maximize the time spent to make each portfolio entry useful.

“Each portfolio entry consists of several questions to guide the learner through the exercise, an evaluation of the exercise and a self-assessment of the skills needed for completion of each exercise. After completing each exercise, the residents reviews their work with a faculty member.”

The resident portfolio includes several self-directed intensive care unit learning opportunities – from obtaining and applying data at the bedside for therapeutic intervention to reflecting on interactions with families. Exercises were developed for each of the core competencies, selecting individual topics related to curricular requirements for critical care medicine defined in the ACGME program requirements. Each portfolio entry consists of several questions to guide the learner through the exercise, an evaluation of the exercise and a self-assessment of the skills needed for completion of each exercise. After completing each exercise, the residents reviews their work with a faculty member. Examples of portfolio entries include: waveform analysis from a pulmonary artery catheter, arterial blood gas analysis with manipulation of ventilator

settings, total parenteral nutrition prescription, and an interview with family members in the ICU. Four of the portfolio entries, pro-con debates, morbidity and mortality conferences, ethics cases, and journal clubs, are presented publicly at a weekly multidisciplinary critical care conference. These portfolio entries, along with the discussion that ensues between participants at the conference, are then posted on our website. In aggregate, the website, containing actual portfolio entries and other learning resources, has become our program portfolio.

We believe that accurate self-assessment is an important

“Our new curriculum allows residents the opportunity to practice self-assessment in a predictive, concurrent and summative fashion – skills that are essential for self-directed, lifelong learning.”

skill which must be fostered during residency. By creating portfolio entries that prompt both self-reflection and assessment and by providing residents with data on their performance, as witnessed by multiple observers and with different assessment tools, we can begin to demonstrate resident competence while simultaneously offering feedback on self-assessment. Our new curriculum allows residents the opportunity to practice self-assessment in a predictive, concurrent and summative fashion – skills that are essential for self-directed, lifelong learning.⁸ ■

Alison Clay, MD, is an assistant professor in Critical Care Medicine in the Departments of Surgery and Medicine at Duke University Medical Center, Durham, NC.

¹ Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad. Med.* 1991; 66(12): 762-769.

² Ward M., Gruppen L, and Regehr G. Measuring Self-assessment. *Adv. Health Sci Educ* 2002; 7(1): 68-80

³ Handfield-Jones RS, Mann KV, Challis ME, Hobma SO, Klass DJ, McManus IC, Paget NS, Parboosingh IJ, Wade WB, Wilkinson TJ. Linking assessment to learning: a new route to quality assurance in medical practice. *Med. Educ.* 2002; 36(10): 949-958.

⁴ Gordon MJ. Self-assessment programs and their implications for health professions training. *Acad. Med.* 1992; 67(10): 672-679.

⁵ Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad. Med.* 2004; 79 (10 Suppl): S70-81.

⁶ Clay, AS, Que, L, Petrusa, E, Sebastian, M. and Govert, J. 360 Degree Assessment in the Intensive Care Unit. Submitted for publication.

⁷ Clay, AS, Que, L, Petrusa, E, Govert, J, and Sebastian M. Debriefing in the ICU: a tool for bedside teaching. Submitted for publication.

⁸ Eva, KW and Regehr G. Self-assessment in the Health Professions: a reformulation and research agenda. *Acad. Med.* 2005; 80(10 Suppl): S46-S54.

Assessing the ACGME Competencies with Methods That Improve the Quality of Evidence and Adequacy of Sampling

Reed G. Williams, PhD, Gary L. Dunnington, MD

“No human being is constituted to know the truth, the whole truth and nothing but the truth; and even the best of men must be content with fragments, with partial glimpses, never the full fruition” *Sir William Osler, p.398¹*

Much of the research on clinical performance assessment has focused on testing-format selection, a surface feature of the assessment process. Van der Vleuten and Schuwirth² have made a strong case that this emphasis has been misplaced. Specifically they demonstrate that testing format has little to do with reliability of assessments. **Table 1** is a modified version of a table from their paper. As can be seen, reliabilities for two hours of testing time vary from 0.53 to 0.84. Further, the subjectively scored testing formats (oral examinations, mini-CEX, and global ratings) are as reliable as the objectively scored formats.

These results make it clear we need to look elsewhere for ways to improve clinical performance appraisal. Research over 30 years has demonstrated that one key to reliable clinical performance assessment is better sampling of the domain.³ Enough examples of performance should be measured to insure that the resident can perform each task consistently. Further, the measurements should systematically cover all task variations and performance conditions.

“Enough examples of performance should be measured to insure that the resident can perform each task consistently. Further, the measurements should systematically cover all task variations and performance conditions.”

In addition to sampling, valid and reliable assessment of ACGME competencies depends on the adequacy of activities and operations (types of evidence) used to measure each. The types of evidence used in global ratings of ACGME competencies include direct personal observation (observing a resident: taking a history, performing procedures, educating patients), indirect evidence (e.g., inferring adequacy of data collection based on observing a verbal or written report of

Table 1

Reliability estimates of different assessment instruments as a function of testing time*

Instrument	Description	Reliability for Two Testing Times	
		2 hours	4 hours
Multiple choice ¹¹	Short stem and short menu of options	0.76	0.93
Patient management problem ¹¹	Simulation of patient, full scenarios	0.53	0.69
Oral examination ¹²	Oral examination based on patient cases	0.69	0.82
OSCE ¹³	Simulated realistic clinical encounters in round robin format	0.69	0.82
Global ratings of clinical performance ¹⁰	Periodic appraisal of clinical performance by faculty members based on observation over a one month period	0.84	0.91
Global ratings of professional behavior ¹⁰	Periodic appraisal of professional behavior by faculty members based on observation over a one-month period	0.81	0.89
Mini-clinical exercise (mini-CEX) ⁵	Evaluation based on observed real patient encounter and short follow-up oral examination	0.84	0.92
Incognito standardized patients ¹⁴	Real consultations scored by undetected simulated patients	0.76	0.82

*Adapted from van der Vleuten and Schuwirth² with additions based on Williams et. al.¹⁰ Reprinted with permission.

findings, inferring communication skills with patients based on case presentation skills or communications with attendings) and hearsay evidence (e.g., evaluating relations with health care personnel based on anecdotal reports from nurses). Hearsay evidence is second hand-information based on something someone else said rather than on what has been seen or experienced personally. Further, the quality is often unknowable. Was the anecdotal report based on a single

of rotations. Second, performance observations are made while the observer functions as a member of the surgical team, diluting attention directed toward assessment.

Table 2 lists representative ACGME competencies and provides our impression of the types of evidence commonly used for assessing each. The table also indicates the frequency and thoroughness of the competency's measurement. By our accounting of the 28 ACGME competencies, 6 are typically supported by direct observation, 18 by indirect evidence and four by hearsay. Twenty-two competencies are infrequently measured and 26 are measured unsystematically. If you accept this analysis, you may agree that simply adding items to existing global rating forms is unlikely to provide a solid basis for judging the ACGME competencies.

For these reasons one SIU Surgery Department assessment priority has been to identify competencies where the quality of evidence and the adequacy of sampling are weakest and to develop and implement assessment procedures that improve on the situation. The remainder of this article describes three of those efforts.

Operative Performance Rating System-based on Sentinel Cases

When it comes to observation, surgeons have an advantage as they stand across from residents in the operating room and watch them perform for hours at a time. However, surgery departments have not taken fullest advantage of this opportunity. Surgeons rarely fill out resident performance rating forms during or immediately after surgery. Therefore the recorded ratings are affected by memory loss and selective recall. In addition, the forms used for evaluating surgical performance have been all-purpose forms.

“Resident assessment based on direct observation often has two weaknesses. First, observers do not record their observations and their judgments at the time of the observation. Ratings are filled out at the end of rotations. Second, performance observations are made while the observer functions as a member of the surgical team, diluting attention directed toward assessment.”

observed incident or on multiple incidents? Hearsay evidence is prone to the problems associated with anecdotal evidence (e.g., extreme and recent events get undue attention).

There is a tendency to equate personal observation data with high quality measurement. However, resident assessment based on direct observation often has two weaknesses. First, observers do not record their observations and their judgments at the time of the observation. Ratings are filled out at the end

Table 2

Quality of evidence used in evaluating selected ACGME competencies

ACGME Competency	Types of Evidence Used Most Frequently to Assess This Competency	Sampling (Frequency, Systematic or Not)	Comments
Patient Care – gather essential and accurate information about patients	Indirect	Infrequent, Not Systematic	Act of taking a history or performing a physical examination is rarely observed
Patient Care – counsel and educate patients and their families	Hearsay Indirect	Infrequent, Not systematic	This competency is rarely observed. Patient satisfaction data is rarely collected systematically
Patient Care – perform competently all medical and invasive procedures essential for the area of practice	Direct Observation	Infrequent, variable, Not systematic	Varies from specialty to specialty and from procedure to procedure
Medical Knowledge – know and apply the basic and clinical sciences appropriate to their discipline	Direct Measurement	Annually, Systematic	In training examinations provide broad and systematic sampling. Clinical impressions of knowledge are frequent but non systematic
Practice Based Learning – analyze practice experience and perform practice-based improvement activities using a systematic methodology	Indirect	Infrequent, Not Systematic	Some programs are instituting activities designed to teach and assess this competency
Practice Based Learning – locate, appraise, and assimilate evidence from scientific studies related to their patients' health problems	Indirect	Infrequent, Not systematic	
Interpersonal and Communication Skills – create and sustain a therapeutic and ethically sound relationship with patients	Hearsay	Infrequent, Not systematic	
Interpersonal and Communication Skills – work effectively with others as a member or leader of a health care team or other professional group	Direct Observation Hearsay	Regular, Not systematic	
Professionalism – demonstrate respect, compassion, and integrity; responsiveness to the needs of patients and society that supersedes self-interest; accountability to patients, society, and the profession; and a commitment to excellence and on-going professional development	Indirect	Infrequent, Not systematic	
Professionalism – demonstrate a commitment to ethical principles pertaining to provision or withholding of clinical care, confidentiality of patient information, informed consent, and business practices	Indirect Hearsay	Infrequent, Not systematic	
Systems-Based Practice – practice cost-effective health care and resource allocation that does not compromise quality of care	Indirect	Rare, Not systematic	
Systems-Based Practice – advocate for quality patient care and assist patients in dealing with system complexities	Indirect	Rare, Not Systematic	

Our primary goal in developing the operative performance rating system (OPRS)⁴ was to optimize the advantages of close contact in the operating room by doing two things. First, we wanted to minimize the elapsed time between observation of surgical performance and recording of judgments about that performance. Second, we wanted to tailor rating forms to operative procedures. We realized it was not practical to have all resident performance observed and assessed, so we selected two sentinel operative procedures for each post graduate year. A ten-item evaluation instrument unique to each sentinel procedure was created. Items focused on procedure-specific operative skills and operative decision-making. Four general items, developed at the University of Toronto, covering tissue handling, time and motion, flow of operation, and overall performance also were included on each instrument. The goals were to: encourage attending surgeons to complete the operative performance rating instrument as the resident performed or immediately afterward; encourage attending surgeons to discuss the performance with the resident shortly afterwards while the performance and the circumstances were fresh in the minds of both parties; and assure that all residents mastered each of these sentinel procedures. Residents are certified after completing the procedure at least three times (minimum of two different raters) with no item ratings below four (five point scale). All residents are evaluated as many times as necessary to meet proficiency certification standards. This system has instructional, motivational and quality assurance benefits.

While the OPRS system applies only to surgical specialties, there are opportunities for other specialties to modify their assessment practices and achieve similar benefits. For example, the American Board of Internal Medicine has designed the mini-CEX for evaluating single resident-patient encounters as a complement to the traditional global rating process.⁵

Patient Assessment and Management Examination (PAME)

MacRae and her colleagues developed the Patient Assessment and Management examination (PAME) process to address key aspects of clinical competence not well assessed using traditional assessment methods.⁶ We have adopted and used the PAME assessment process with our PGY 3-5 residents for the past five years. The resident first interviews and examines the patient. Based on the information collected, the resident educates the patient and suggests what additional diagnostic tests are needed to confirm the diagnosis. After ordering diagnostic tests and studying the results in another room, the resident returns to see the patient, explains the results, recommends management, explains options, and answers questions the patient has about the test results and management plans. Both encounters are observed by an

attending surgeon through one-way glass. The surgeon and the resident go to a conference room where the attending surgeon asks questions designed to better understand the underlying basis for the resident's actions, and recommendations. The primary benefits of the PAME assessment system are that the resident: 1) makes all decisions independently, 2) is responsible for all facets of the two encounters, 3) is observed throughout the process of interviewing, examining and educating the patient, 4) must explain her/his decisions and actions during the oral examination portion. Attending surgeons have accepted this assessment method well because it provides them with access to facets of resident performance that they normally do not see. Further, the examination can be scheduled six months in advance allowing them to incorporate it into their schedules with minimal interference.

“The primary benefits of the PAME assessment system are that the resident: 1) makes all decisions independently, 2) is responsible for all facets of the two encounters, 3) is observed throughout the process of interviewing, examining and educating the patient, 4) must explain her/his decisions and actions during the oral examination portion.”

Progress decisions by committee

There are ways to bolster the value of global resident performance ratings and faculty confidence in the process. In a five year study, we learned that 18% of significant resident performance deficiencies (those the faculty required residents to remediate) only became apparent when attending physicians came together to discuss performance at the annual evaluation meeting.⁷ These performance deficiencies had not been raised by any faculty member on any end-of-rotation evaluation form at any other time during the year.

Annual resident evaluation meetings can help in at least two ways. They can 1) make patterns of resident behavior apparent that were not apparent to individual physicians, and 2) provide evidence that strengthens individual surgeon pre-existing convictions about residents' performance deficiencies.

The SIU general surgery program resident evaluation meetings are held twice a year. All faculty members are invited to attend. Those who attend are provided with a portfolio including all performance data available for each resident (summaries of all end-of-rotation global ratings and comments submitted by faculty members, annual and mid-year summaries of performance by the resident's advisor, and

written performance summaries documenting resident performance and progress decisions from prior years). The portfolios vary from 6–13 pages per resident with longer portfolios belonging to more senior residents and those with a cumulative history of performance deficiencies. Discussions of individual residents average about 5 minutes in length with a range from 1 minute to 30 minutes. Most discussion time is focused on residents with performance problems.

“There was no evidence of piling on or feeding frenzies during the meeting. Participants felt that their comments were more anonymous in the meeting thus freeing them to raise issues (e.g., resident honesty) that they did not raise under other conditions.”

While our study⁷ and those of others⁸ suggest that resident progress decisions by committee provide a broader perspective on performance, and detect problems that would otherwise go unnoted, there has been countervailing concern that group dynamics may compromise the quality of progress decisions. Many people fear that progress committee meetings may be dominated and unduly influenced by strongly opinionated or powerful individuals. In a study to determine the effects of group dynamics on resident progress decisions,⁹ we found no evidence that led us to believe individual participants dominated discussion or had undue influence by their persuasive powers or their position in the organization. Likewise, there was no evidence of piling on or feeding frenzies during the meeting. Participants felt that their comments were more anonymous in the meeting thus freeing them to raise issues (e.g., resident honesty) that they did not raise under other conditions.

We believe that these three changes have increased direct observation of important clinical competencies, assured us of more systematic and focused appraisal and have provided better triangulation on the performance of our residents. In this way the changes have complemented our existing global performance rating process. ■

Reed G. Williams, PhD, is Professor and Vice Chair for Educational Affairs and Gary L. Dunnington, MD, is Professor and Chair, Department of Surgery, Southern Illinois University School of Medicine, Springfield, Illinois.

- ¹ Osler W. *Aequanimitas*. Third ed. Philadelphia: P. Blakiston's Son & Co., Inc., 1932.
- ² van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med. Educ.* 2005; 39(3): 309-17.
- ³ Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical competence ratings. *Teach. Learn. Med.* 2003; 15(4): 270-92.
- ⁴ Larson JL, Williams RG, Ketchum J, Boehler, ML, Dunnington, GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgical residents. *Surgery* 2005; 138(4): 640-9.
- ⁵ Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann. Intern. Med.* 2003; 138(6): 476-81.
- ⁶ MacRae HM, Cohen R, Regehr G, Reznick, R, Burnstein, M. A new assessment tool: the patient assessment and management examination. *Surgery* 1997; 122(2): 335-43; discussion 343-4.
- ⁷ Schwind CJ, Williams RG, Boehler ML, Dunnington GL. Do individual attending post-rotation performance ratings detect resident clinical performance deficiencies? *Acad. Med.* 2004; 79(5): 453-457.
- ⁸ Hemmer PA, Hawkins R, Jackson JL, Pangaro LN. Assessing how well three evaluation methods detect deficiencies in medical students' professionalism in two settings of an internal medicine clerkship. *Acad. Med.* 2000; 75(2): 167-73.
- ⁹ Williams RG, Schwind CJ, Dunnington GL, Fortune, J, Rogers, D, Boehler, M. The effects of group dynamics on resident progress committee deliberations. *Teach. Learn. Med.* 2005; 17(2): 96-100.
- ¹⁰ Williams RG, Verhulst SJ, Colliver JA, Dunnington GL. Assuring the reliability of resident performance appraisals: More items or more observations? *Surgery* 2005; 137(2): 141-7.
- ¹¹ Norcini JJ, Swanson DB, Grosso LJ, Webster GD. Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence. *Med. Educ.* 1985; 19(3): 238-47.
- ¹² Swanson DB. A measurement framework for performance-based tests. In Hart IR, Harden R, eds. *Further Developments in Assessing Clinical Competence*. Montreal: Can-Heal publications, 1987. pp. 13-45.
- ¹³ van der Vleuten CP, van Luyk SJ, Swanson DB. Reliability (generalizability) of the Maastricht Skills Test. *Proc. Annu. Conf. Res. Med. Educ.* 1988; 27: 228-33.
- ¹⁴ Gorter S, Rethans JJ, van der Heijde D, Scherpbier, A, Houben, H, van der Vleuten, C, van der Linden, S. Reproducibility of clinical performance assessment in practice using incognito standardized patients. *Med. Educ.* 2002; 36(9): 827-32.

Planning for Remediation

Scott A. Schartel, DO

Advancement through the educational process is simple for the student who consistently meets expectations, achieves expected goals, and passes all assessments. However, not all students will achieve expected milestones at the same rate; some will never achieve acceptable performance. The challenge for educators is to help the student who is slow to meet expectations to meet them in a reasonable period of time and to identify those students who are unlikely to ever reach an acceptable level.

“Not all students will achieve expected milestones at the same rate; some will never achieve acceptable performance. The challenge for educators is to help the student who is slow to meet expectations to meet them in a reasonable period of time and to identify those students who are unlikely to ever reach an acceptable level.”

In the Oxford English Dictionary remediation is defined as “the action of remedying....”¹ Among the definitions for remedy is, “... to put right, reform (a state of things); to rectify, make good.” This definition stresses that the fundamental purpose of educational remediation is to help the student to succeed. Successful remediation requires that the teacher identify those areas of performance that are deficient. Clear goals and objectives, matched with an evaluation system that can identify a student’s achievement of the stated goals and objectives, will allow the teacher to formulate a problem list. After reviewing the problems identified, the teacher can group problems into broad categories. Useful categories could be the three educational domains: knowledge, skills, and behaviors. Alternatively, the ACGME general competencies can be used as an organizing framework. By using an organizing principle to categorize the student’s problems, a coherent plan for remediation can be established.

Having identified the need for remedial education, the teacher must meet with the student to discuss the areas of deficiency. The more specific this discussion, the more useful it will be to the student. To tell a resident that “the faculty thinks you are lazy” does not help the resident improve. It is better to tell the resident: “The faculty have observed that you are frequently late for clinic, your notes are incomplete, you do not complete dictations until you have been threatened with disciplinary action, and nurses report that you are not responsive to requests to see patients.”

After the problem areas are identified to the resident, the resident should be given a plan for corrective action. If the areas of deficiency are behavioral, then specific expected

behaviors should be identified (e.g., arrive on time for clinic, complete all dictations by the initial deadline, answer all requests to see patients when on-call within 10 minutes, etc.). For skill deficiencies, the plan may include additional time assigned to rotations or activities that allow the resident to practice the skill. It may include learning in a skills or simulation laboratory. Use of a preceptor or a limited number of faculty physicians to supervise and teach the resident during the remedial period offers the opportunity for a consistent educational approach and consistent feedback.

When deficiencies are primarily in the cognitive domain, an attempt should be made to further define the resident’s areas of weakness and to identify contributing factors. A poor knowledge base may be related to poor study habits, a lack of effort, or a learning disability. Questioning the resident about what, where, when, and for how long s/he is studying may identify poor study habits or insufficient effort. It may be useful to ask the resident to keep a study diary. Identifying that the resident reads two to three nights per week, sitting in bed at the end of the day, usually falling asleep within a few minutes provides an opportunity for the teacher to help the resident to recognize his/her poor study skills and make suggestions for correction.

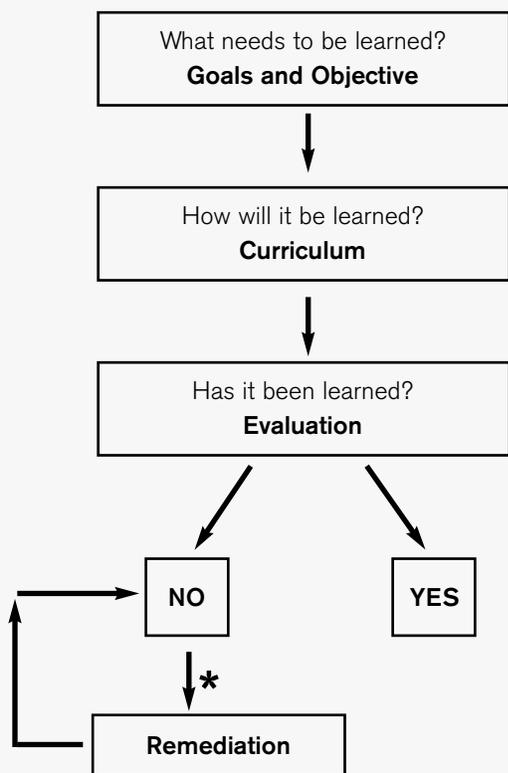
Failure in the knowledge domain, especially if manifested by poor performance on standardized tests, may be related to a learning disability. While some medical educators have been skeptical – questioning whether students with learning disabilities can reach the level of medical school or residency – there is data to suggest this does occur.² Typically these will be residents who succeed, or even excel, in most aspects of their

“While some medical educators have been skeptical – questioning whether students with learning disabilities can reach the level of medical school or residency – there is data to suggest this does occur.”²

education, except for their performance on standardized multiple choice examinations. Rosebraugh, in his discussion of learning disabilities in medical students, reports that these students often have IQs in the above average to gifted range, but often have a history of difficulty with standardized tests, especially in the area of reading comprehension. He observes that for these individuals the problems may not be discovered until medical school or later when they are no longer able to compensate due to the increased demands of the curriculum.² Thoughtful discussions of the issue of learning disabilities in medical education can be found in the essays of Hafferty and Gibson.^{3,4}

With remedial education the resident who is failing in the knowledge domain may be able to learn strategies to deal with her/his learning disabilities and achieve success on examinations. The Medical H.E.L.P program at Marshall

Figure 1
Learning, Evaluation, and Remediation



* After the first round of remediation, a decision must be made about further remediation vs. dismissal.

University is an example of a program designed to identify and help medical students and graduate physicians with learning disabilities.⁵

After the resident's deficiencies are identified and remediation goals are established, the resident should be clearly told what objectives must be met by the end of the remedial period in order to achieve a satisfactory evaluation. The resident must be given specific details about how s/he will be evaluated, by whom, and at what frequency during the remedial period. It should be made clear to the resident what the consequences of successful or unsuccessful remediation will be. If it is necessary for a resident to repeat a rotation, the resident must be informed about how this will effect the duration of training. During the remediation, the frequency of evaluation and feedback should generally be increased. This will help both the resident and program director to judge progress. It will also allow for timely modification of the remediation plan if it is not working.

If the resident has not achieved an acceptable level of performance at the end of the remediation period, decisions must be made about what, if any, progress has been achieved and if the resident will benefit from additional remediation.

These can be very difficult decisions. If the resident has made progress, the rate of progress and the level of remaining deficiency may help answer the question about the benefit of further remediation. The use of multiple evaluators can help the faculty to reach a broad-based consensus.

If a consensus is reached that a resident will not succeed, even with additional remediation, then the resident should be informed of this conclusion. The resident deserves to be given information to explain why this decision was made. While dismissal from a program is a serious step, and should not be made without due consideration, allowing a resident to continue when the faculty have concluded that the resident will not succeed is unfair to the resident. It is in the resident's best interest to help her/him to understand and accept the situation. It may help the resident to deal with this traumatic event if s/he is told that the failure to succeed in a particular specialty area does not imply that the resident is a bad or unworthy person, but rather that the resident's aptitudes do not match those required for the specialty. The faculty should try to help the resident understand that the lack of aptitude in one specialty does not preclude the aptitude to succeed in another. Those residents with the most insight will be able to see this easily, while others may require more counseling.

“It is in the resident’s best interest to help her/him to understand and accept the situation. It may help the resident to deal with this traumatic event if s/he is told that the failure to succeed in a particular specialty area does not imply that the resident is a bad or unworthy person, but rather that the resident’s aptitudes do not match those required for the specialty.”

If a decision to dismiss a resident is made, the resident is entitled to due process. The process will be determined by contractual obligations, institutional policy, and state and federal law. It is important that these requirements be followed. While litigation may result even when all the correct procedures have been followed, in general, courts have shown deference to the decisions of faculty about issues related to clinical competence.^{6,7} Courts are most likely to intervene on matters involving the breach of contractual or statutory obligations.

Program directors and faculty have dual obligations when working to remediate the performance of a resident who is failing: an obligation to the resident and an obligation to society. The obligation to the resident stems from the implicit understanding that when a resident is admitted into a residency program, the program faculty will make every effort to help the

“Faculty must always do what they believe is right, putting the interests of patients first. While it may be uncomfortable, unpleasant, and personally distressing to dismiss a resident, program or personal concerns can not supersede the safety and best interests of patients.”

resident succeed. This includes the expectation that faculty will provide residents with constructive formative evaluation in order to maximize the residents’ chances for success.

The faculty has an additional obligation to society to ensure that all residents who complete a graduate medical education program meet the standards for safe and competent practice. This obligation emphasizes the importance of effective summative evaluation. Faculty must always do what they believe is right, putting the interests of patients first. While it may be uncomfortable, unpleasant, and personally distressing to dismiss a resident, program or personal concerns can not supersede the safety and best interests of patients. No physician wants to harm patients or put them at risk, including physicians in training. Therefore, it is in the resident’s best interests not to be allowed to practice in an area in which faculty have determined the resident is incompetent.

Achievement of the best results for all those involved—residents, faculty, and patients – requires a well organized system of evaluation, feedback, and remediation (**Figure 1**). During medical school and residency, physicians do not receive much, if any, education about being an effective teacher or evaluator. It is incumbent upon the leaders of graduate medical education programs to provide this education to their faculty to ensure that evaluation and feedback are provided in an appropriate and effective manner. ■

Scott A. Schartel, DO, is Professor, Associate Chair for Education and Program Director in the Department of Anesthesiology, Temple University, Philadelphia, PA.

¹ Oxford English Dictionary (Online). Quoted definitions from 2nd Ed, 1989. Available at <http://dictionary.oed.com>. Accessed 3/9/06.

² Rosebraugh CJ. Learning disabilities and medical schools. *Med Educ.* 2000; 34: 994-1000.

³ Hafferty FW, Gibson GG. Learning disabilities and the meaning of medical education. *Acad. Med.* 2001; 76:1027-1031.

⁴ Hafferty FW, Gibson GG. Learning disabilities, professionalism, and the practice of medical education. *Acad. Med.* 2003; 78:189-201.

⁵ Marshall University H.E.L.P. Program. <http://www.marshall.edu/help/services/medical/> Accessed 3/10/06.

⁶ Irby DM and Milam S. The legal context for evaluating and dismissing medical students and residents. *Acad. Med.* 1989; 64(11): 639-643.

⁷ Helms LB; Helms CM. Forty years of litigation involving medical students and their education: I. General educational issues. *Acad. Med* 1991; 66:1-7.

Reluctance to Fail Poorly Performing Residents – Explanations and Potential Solutions

*Nancy Dudek, MD, MEd, Meridith Marks, MD, MEd
Glenn Regehr, PhD*

Despite its importance for evaluating the clinical competence of residents, some have questioned the utility of in-training evaluation by faculty for identifying residents in difficulty. Efforts to improve the value of faculty evaluations have mostly focused on developing better rating instruments or on training faculty to use these instruments more accurately (e.g.: teaching faculty that a score of 3/5 corresponds to a particularly defined level of performance). Despite these efforts, the failure to report poor clinical performance continues to be a problem in resident education.

“Our results confirmed that the problem has little or nothing to do with the technical aspects of the scales themselves or with a faculty member’s ability to use them. Rather the problem lies primarily with a faculty member’s willingness to use them.”

In our paper, “Failure to Fail – The Perspectives of Clinical Supervisors,” we suggested that the problem may have less to do with the technical aspects of evaluation than with its social aspects.¹ Prior to our study, there had been little consideration in the medical education literature of the political and relational issues that likely influence how supervisors rate a certain performance. In response to this gap, and in an attempt to determine why clinical faculty do not report poor clinical performance on evaluations, we conducted a qualitative study using semi-structured interviews with clinical supervisors involved in the evaluation of residents. Our results confirmed that the problem has little or nothing to do with the technical aspects of the scales themselves or with a faculty member’s ability to use them. Rather the problem lies primarily with a faculty member’s willingness to use them.

On the positive side, we found that most of the faculty interviewed felt very confident in their ability to determine whether a resident was performing poorly. Further, several did report having failed a student in the past, and the main

motivation expressed for this decision was a sense of responsibility to three sources:

- A responsibility to the public to ensure safety;
- A responsibility to the profession to protect its reputation;
- A responsibility to the resident to allow them the opportunity for remediation.

Despite the acknowledged importance of failing poorly performing residents and, despite faculty members' confidence that they are able to identify the individuals that should fail, many felt that failures were not recorded as often as they should be. In fact, several clinical supervisors admitted to having passed a resident about whom they had concerns. When asked why they might have let this happen, the faculty identified three broad areas that interfered with their willingness to fail students and residents:

- Documentation issues (the lack of appropriate and formal "proof")
- Consequences for themselves (both the additional commitment of time and the public questioning of the validity of their evaluation)
- Consequences for the resident (the lack of an appropriate process for redressing a failing performance)

We will address each of these areas briefly then discuss the implications for future efforts to improve clinical evaluation.

“In fact, several clinical supervisors admitted to having passed a resident about whom they had concerns. When asked why they might have let this happen, the faculty identified three broad areas that interfered with their willingness to fail students and residents.”

Documentation issues

It has been suggested by some (apparently frustrated) researchers and developers that faculty are essentially untrainable, that they are unwilling to put the time into learning how to evaluate well or to actually watch their residents to know how well they are performing. However, when we asked our participants, they suggested not so much an unwillingness to invest the time or energy, but rather a lack of structures and knowledge to enable this behavior. Supervisors frequently identified having concerns about residents or fellows but were unsure what would constitute “valid evidence” to support that concern. Some suggested that it would be a less onerous task if the supervisor would not have to reconstruct specific incidences of poor resident performance at the end of the clinical rotation. They suggested that technologies may be able to facilitate ongoing documentation of specific examples

“If a resident had received positive evaluations previously, the supervisor was reluctant to fail him/her, especially if they respected the person/people who had made the previous judgment.”

of performance, but that knowing what type of information to document would be more important in failing a resident than the existence of the technology itself. Thus, the faculty development issue is not one of how to use the scale properly at the end of the rotation, but what specifically needs to be documented as “proof” and how to document that information in a practical manner when concerns start to arise during the rotation.

Consequences for faculty

The supervisors in our study also seemed dissuaded from failing a student or resident by the potential for consequences for themselves, especially in the context of an appeal process. For example, those who had gone through the process indicated a sense of isolation when an evaluation they had provided was appealed. They agreed that residents needed to be given the opportunity to appeal their grades but felt that the supervisors were largely unsupported during the process. In addition, participants were concerned about having their credibility threatened during a formal appeal process. This finding is interesting given that participants told us that they feel confident in their ability to identify residents and fellows who are performing poorly. So why do they lose confidence when they are challenged?

Often the student or resident has received positive evaluations in the past, and other research has suggested that being the first to “blow the whistle” is difficult. Instead, people will go against their own perception of the truth to conform with the group.² Additionally, “informational social influence”, whereby others are seen as a source of information to guide a person when the situation is ambiguous, may also be contributing.² A number of supervisors admitted that when they had a poorly performing resident they would attempt to find out how the individual had been evaluated in the past. If a resident had received positive evaluations previously, the supervisor was reluctant to fail him/her, especially if they respected the person/people who had made the previous judgment. This situation may be exacerbated by another social phenomenon, the “diffusion of responsibility.” Studies have demonstrated that the greater the number of witnesses to a situation, the less likely it is that someone will intervene.² In the case of the poorly performing individual, the supervisor may feel less responsibility because they know that several other supervisors will evaluate that person during their residency.

Consequences for the resident

Supervisors also indicated concern about the resident's reputation and the lack of a system in place for remediating a failing resident. The impact of a single evaluation on the individual's overall program evaluation appeared to be overestimated by the study participants. In addition, the faculty tended to assume that no remediation was available (which may not always have been correct). The sense of obligation to the resident may in part be reflective of the supervisors' sense of obligation to the learner as a protégé, but also seemed to reflect a lack of knowledge both of the evaluation process and of whose responsibility it is to arrange appropriate remediation. Many were unwilling to fail a resident unless a good remediation option was apparent to them.

Implications for the improvement of in-training faculty evaluation ratings

Our findings lead us to look in very different places for improving the utility of the in-training faculty evaluations. Rather than trying to improve the scales themselves or teach faculty how to use them, our results suggest that we should be addressing the systemic, social, psychological and relational issues that impede faculty from assigning a failing grade.

From a systemic perspective, the institution clearly needs to establish mechanisms to support the assignment of a failing grade. For example, a resource office and support system might be provided for clinical supervisors. Such a resource would serve many roles. First, it would act as a point of contact for supervisors when they first recognize that they are dealing with a learner who is failing to meet expectations. This office would clarify what information needs to be collected and

“From a systemic perspective, the institution clearly needs to establish mechanisms to support the assignment of a failing grade. For example, a resource office and support system might be provided for clinical supervisors.”

what steps need to be taken to comply with the university's evaluation and appeals process. Second, this office would counsel the supervisor on how to provide “bad news” to residents, present information in a written evaluation to support their impressions, how to interact with the resident who challenges the supervisor's opinion, and how to handle the appeals process, if it occurs. Third, the office would provide support to the supervisor who is going through an appeal. If no institutional resource office exists, all of these roles could be provided by the residency program director and/or member of the residency program committee.

The program director needs to also make it clear to the clinical supervisors that creating a remediation program for the failing resident is not their responsibility. Rather, the program director and members of the residency program's committee are responsible for this. Once clinical supervisors become aware that remediation programs for residents are possible they may be more willing to fail the resident so that they can get the assistance needed to improve their performance.

“Program directors also must find ways to emphasize the importance of identifying failing residents who are not meeting expectations early in their training, when there is time left to remediate them. They must also encourage the individual supervisor's sense of responsibility to report what they have judged to be a failing performance.”

From a more personal perspective, there are several issues that must be addressed to encourage faculty to formally record a negative evaluation of a resident. First, there appears to be a diffusion of responsibility phenomenon, which contributes to the problem. A supervisor knows that the resident will have more training before entering independent practice. Thus, it is often easy to make this “somebody else's problem.” Program directors also must find ways to emphasize the importance of identifying failing residents who are not meeting expectations early in their training, when there is time left to remediate them. They must also encourage the individual supervisor's sense of responsibility to report what they have judged to be a failing performance. Second, there also appears to be some pressure to conform to the interpretation of others that this individual is meeting expectations. Program directors must find ways to emphasize the value of each individual supervisor's judgment even if it is not consistent with the opinion of others. If program directors let supervisors know that concerns have been raised in the past, they may feel more comfortable in being the first person to fail that resident or fellow. These are undoubtedly difficult challenges for program directors, but success may be in appealing to the supervisors' sense of responsibility to the public, the profession and the resident.

Conclusions

The willingness of supervisors to identify failing residents is an extremely important area of medical evaluation to explore further. In-training faculty evaluations are an important component of resident evaluations and are here to stay. In many situations, especially the complex clinical encounters,

“In fact, several clinical supervisors admitted to having passed a resident about whom they had concerns. When asked why they might have let this happen, the faculty identified three broad areas that interfered with their willingness to fail students and residents.”

we do not have another means of evaluation available. Understanding the social, psychological and systematic pressures that stop supervisors from expressing the true level of ability that they have determined in their clinical residents is necessary if we are going to develop an evaluation system that enables clinical supervisors to consistently report poor clinical performance. It appears possible to make this process easier for supervisors by providing them with more information and support when faced with a failing resident. ■

Nancy Dudek, MD, MEd, is Assistant Professor, Faculty of Medicine, Division of Physical Medicine & Rehabilitation, University of Ottawa. Meredith Marks, MD, MEd, is Assistant Dean for Professional Affairs, Faculty of Medicine, University of Ottawa; Glenn Regehr, PhD, is the Richard and Elizabeth Currie Chair in Health Professions Education Research and Professor, Faculty of Medicine, University of Toronto.

¹ Dudek NL, Marks MB & Regehr G. Failure to fail – the perspectives of clinical supervisors. *Acad. Med.* 2005; 80(10):S84-7.

² Ross L, Nisbett RE. *The person and the situation – perspectives of social psychology.* Temple University Press, Philadelphia, 1991.

A Multi-Faceted Approach to Resident Evaluation and Remediation

Phillip M. Boiselle, MD, Bettina Siewert, MD

Evaluation and remediation of residents are essential components of a residency program. While an effective evaluation system has the potential to detect problems early, an operative remediation program is necessary to effectively deal with issues once they are identified.

In recognition of the importance of early detection and prompt remediation of residents experiencing difficulties, we sought to develop a multifaceted approach to resident evaluation with the aim of early identification and prompt remediation of difficulties. This article briefly describes both the comprehensive evaluation program and the formal remediation program that we have developed and successfully implemented in our radiology residency program.

Multi-source evaluation system

Our former system of resident evaluation placed a disproportionate emphasis upon resident rotation evaluations. Although rotation evaluations continue to comprise an integral component of our revised evaluation system, our new system involves input from several complementary sources, including:

- **regular evaluation and feedback of performance** in each of the 6 competencies;
- **regularly scheduled “roundtable” faculty discussions** to help identify “sub-threshold” but potentially important resident performance issues that may “escape” detection on written evaluations;
- **call setting evaluations** to assess the ability of residents to perform independently;
- **American Board of Radiology in-service written examination scores** to assess general fund-of-knowledge;
- **resident self-assessment** to help highlight learner-identified areas of need; and
- **360-degree evaluation** to enhance assessment of professionalism and interpersonal/communication skills.

Although the benefits of most of these methods are well established, we suspect that the concept of a regularly scheduled “roundtable” faculty discussion will be less familiar to some readers. In our experience, this forum has brought to the forefront several resident issues that were initially not reported on written evaluations because the faculty members thought that the observed performance or behavior was simply due to a “sampling error” or was “within the edge of a bell-

“In our experience, this forum has brought to the forefront several resident issues that were initially not reported on written evaluations because the faculty members thought that the observed performance or behavior was simply due to a ‘sampling error’ or was ‘within the edge of a bell-shaped curve.’”

shaped curve.” Upon discussion with other faculty members, however, it was often demonstrated that such concerns were, in fact, more generalized. Thus, this forum has helped to bring resident difficulties to the attention of the program directors at an earlier stage, at a time when such problems are potentially more responsive to remediation.

Evaluation and remediation

Figure 1 shows the relationship between evaluation and remediation. Both are essential and interrelated components of a residency program. While an effective evaluation system has the potential to detect problems early, an operative remediation program is necessary to effectively deal with issues once they are identified. In turn, an evaluation system is necessary to determine whether remediation has been effective. The resident, program director, and faculty members play important and overlapping roles in both the evaluation and remediation processes.

Consistent remediation framework

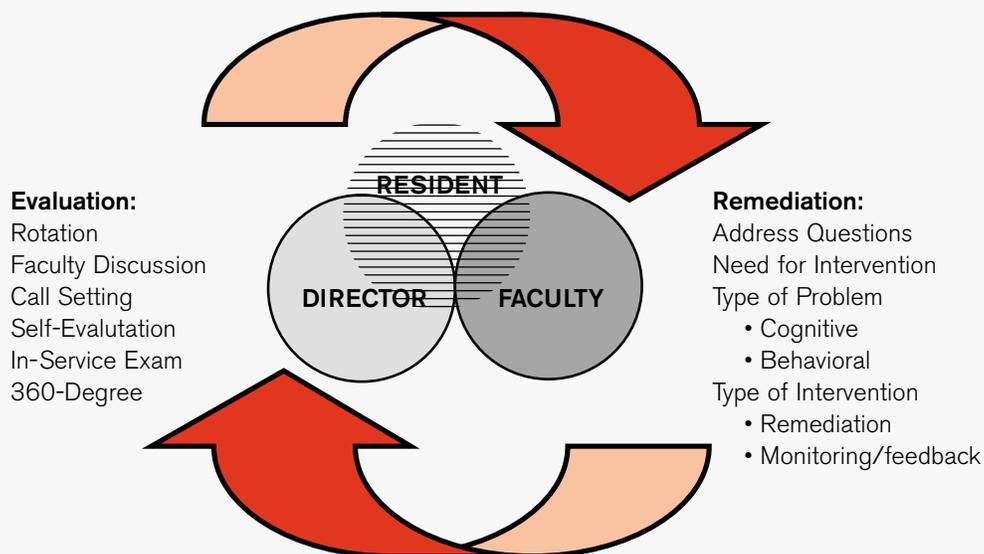
Once a resident performance issue has been identified, it is important to perform a fair and comprehensive assessment in a timely manner, addressing several essential and interrelated questions:

- Is the perceived performance issue primarily related to a knowledge deficit, an attitude or behavioral problem, a skill deficit, or a combination of these factors?
- What is the resident’s perception of the issue?
- What is the faculty’s perception of the issue?
- Are there other contributing personal issues such as depression, anxiety, or personal or family illness?
- Are there larger educational or system problems in the residency program that may contribute to this problem?
- What is the potential impact of the problem with regard to patient care?
- What is the potential impact of the problem in the area of the resident’s personal and professional growth?
- What is the potential impact on other residents and faculty?

Based upon this analysis, a decision should be made as to whether an intervention is necessary. If one is deemed appropriate, it should be put into place with all due deliberate speed.

In order to determine the appropriate type of intervention, it is necessary to categorize the problem as either primarily cognitive (related to the resident’s knowledge base and cognitive skills) or primarily behavioral (related to difficulties

Figure 1
Evaluation and Remediation



[Reprinted from: “A remedy for resident evaluation and remediation,” Academic Radiology, 12(7):894-900, Boiselle PM, © 2005 with permission from Association of University Radiologists]

with professionalism and interpersonal communication) in nature. Cognitive problems are typically amenable to traditional instructional methods. Behavioral issues are more likely to respond to other methods, such as close monitoring/feedback of attitudinal and interpersonal behaviors for “mild” problems and referral to a psychologist or psychiatrist for residents with serious psychological symptoms.

Although each resident’s problems are unique, we sought to develop a framework for intervention that could be applied consistently among residents with various performance issues. The centerpiece is a resident-program director-faculty

“A discernable strength of this framework is that it sets up an alliance between three parties: the resident, the program director, and the faculty educational liaison with experience in the specific area of difficulty.”

educational liaison agreement form that documents the problem and its significance, outlines a detailed remediation and evaluation plan, and communicates the goals of the plan.

A discernable strength of this framework is that it sets up an alliance between three parties: the resident, the program director, and the faculty educational liaison with experience in the specific area of difficulty. With all three invested in the agreement, the “intervention” feels like a concerted team effort rather than a punitive situation and frames the intervention itself in positive rather than oppositional terms. Such a plan may include additional readings, assignment to a faculty mentor, one-on-one tutorials, and repeating a clinical rotation after completion of such measures.

Our remediation program was widely accepted and approved by our departmental educational committee. Both faculty and residents alike expressed approval for having a system in place to address serious performance issues in a consistent and even-handed manner. Importantly, the detailed agreement form documents these issues, which is a critical part of “due process” procedures at any academic institution.

Examples

We provide two fictional examples that illustrate the types of resident difficulties that we have addressed with our remediation program.

Fictional resident John demonstrated an appropriate level of knowledge and clinical skills for his level of residency. At the same time, he was frequently late for work, took personal

cell phone calls during read-out sessions, and occasionally left work before completing his clinical responsibilities. His problem clearly resided in the behavioral realm. Upon discussion of these behaviors with John, it became clear that he had poor insight into his lack of professionalism. Following a frank conversation about the impact and significance of his behaviors and a review of our program’s specific expectations for professionalism, a remediation program was devised. This included assignment to a faculty mentor and close monitoring and feedback of various aspects of professionalism, with specific goals for attendance, punctuality, and overall work ethic. Following successful completion of the remediation program, John now is consistently more professional in his behavior and reports that he feels “more like a real doctor” compared to before the intervention.

In contrast, fictional resident Sally routinely arrived at work early and stayed late. Yet, she struggled on her clinical rotations, lagging behind her peers in terms of her fund of knowledge and her ability to synthesize clinical and radiographic data. Although still in the early stages of residency, she was already involved in several research projects. Upon discussion with Sally, it was learned that she lacked basic time management skills and that she was having difficulty juggling her clinical training with her research interests. A remediation program was devised that included time management education, maintaining a weekly logbook documenting her studying habits, and a hiatus from her research projects until she met the goals set for self-study and improved clinical performance. Upon completion of the remediation program, Sally’s clinical performance is now “on target” and she reports a greater sense of well-being compared to before the intervention. She plans to wait until she has dedicated elective time to return to her research projects.

Although these examples are both success stories, we recognize that there are instances in which a resident may not be able to overcome his or her performance issues despite the implementation of an appropriate plan of action. For such residents, prompt referral for supportive psychological and career counseling is essential to help them find an alternative career path that is better suited to their strengths. ■

Philip Boiselle, MD, is Associate Director and Bettina Siewert, MD, is the Program Director of the Radiology Residency Program at Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

Remediation in Graduate Medical Education: A Point of View from Surgery

Christina G. Rehm MD, Pamela A. Rowland, PhD

The goal of graduate medical education is to develop competent physicians. In order to become a surgeon, this process requires individuals to commit to at least a decade of training. This is willingly accepted because for most, this career is not primarily a way to prosperity, but rather a fulfillment of the need for a purpose in life – to contribute, to make a difference. To reach this goal, it is imperative that teachers of nascent surgeons focus on developing residents’ knowledge and skill to competency levels that meet national standards.

“This is willingly accepted because for most, this career is not primarily a way to prosperity, but rather a fulfillment of the need for a purpose in life – to contribute, to make a difference.”

Progress in the acquisition of competence during residency is measured yearly using the ABSITE (American Board of Surgery In-Training Exam) and standardized individual performance evaluations as required by ACGME. The intent of these evaluations is to identify struggling individuals during training so opportunities for remediation can be provided, perhaps avoiding an unsuccessful attempt in the board certification process. Board certification by the American Board of Surgeons (ABS) requires passing both a qualifying written exam (QE) and a certifying oral exam (CE) and there are strict eligibility limits.¹ While completion of a surgical training program implies that the individual is competent, some may fail to pass because they struggle to demonstrate their competence within the framework of the board certification process.

The process of regaining board eligibility has evolved. We examined the effects of those changes on pass rates of readmitted candidates.² From 1980 to 1985, CME was required to regain eligibility. During that time, only 7% of readmitted candidates passed the QE. In 1985, the requirement changed and from 1986 through 1993 an additional residency year was required. QE pass rates improved (20% of readmitted candidates passed) and 73% of readmitted candidates passed the CE, but these results suggest that simple repetition of an additional residency program year without adjustment for

special needs is not effective. In 1995, the ABS implemented an innovative remediation approach, a 12 month structured education requirement in an ACGME-approved remedial year (RY) surgery program. Improvement in the QE pass rate was substantial (46% of readmitted candidates between 1996 and 2002) but there was little change in the CE pass rate (73% of readmitted candidates between 1996 and 2002).

While the number of candidates who repeatedly fail is relatively small, of those who do, only 14% will complete the remediation process for the QE and just 23% will seek remediation for the CE. The consequences of losing board eligibility can be devastating. Not only is it destructive psychologically, but also it becomes nearly impossible to obtain hospital privileges or malpractice insurance. This means that most of these candidates can no longer practice and are left with no real source of income. Inasmuch as the remedial year has been shown to be a viable option only for a few, in 2003 the ABS introduced an alternative pathway for regaining eligibility, which consists of a self study program followed by an abbreviated written examination. This option allows candidates to continue practicing. Initial results have been disappointing, perhaps due to the lack of effective remediation. But the remedial year remains an option.

We examined those institutions identified by the ABS for having outstanding remedial programs and found that very specific elements were present.² Each program:

- Selects candidates who are successful in practice, respected by peers, aware of their deficiencies, willing to accept help;
- Develops methods to identify the specific individual needs; these are different for each individual and different for the written versus oral boards;
- Selects faculty committed to teaching basic general surgery;
- Encourages the candidates to teach themselves (“one does not know what one can’t teach”);
- Offers opportunities to practice communication and presentation skills.

Correction of knowledge deficits was essential for remediating the QE and self-presentation and communication skills were critical for success on the CE. While candidates who participated in these particular remedial programs did indeed succeed, we suggest that in-training remediation programs are likely to be more effective in the early identification of residents at risk of failure as well as those who will not benefit from remediation. In-training programs permit planning for individual needs and close monitoring of progress and should reduce the occurrence of painful failures in board certification.

“Structured ABSITE remediation is mandatory for all residents and consists of a weekly topic-oriented lecture, followed by a computer based topic summary review and ten multiple choice questions”

In our surgery program at Oregon Health and Sciences University we implemented a remediation program containing the following elements:

- Structured ABSITE remediation is mandatory for all residents and consists of a weekly topic-oriented lecture, followed by a computer based topic summary review and ten multiple choice questions.
- Faculty physicians review all residents at regular intervals specifically to identify individuals at risk. Each such resident is paired with a suitable faculty mentor and regular, inconspicuous meetings are scheduled during regular work hours.
- We developed an organization and structure that supports good study and work habits.
- Faculty offer methods for restoration of self-confidence to those who require assistance.

We have implemented this program gradually, and it is too soon to tell which elements work, but for the past three years all of our graduates have successfully completed their boards. Remediation programs should also identify individuals who, despite intense teaching and mentoring efforts, do not appear suited for surgery. An alternative career path should be recommended earlier rather than later.

In recent years much has been written regarding declining applications for surgical training and interest in a surgical career. Therefore, concern about serious work force issues in the future has been raised.³ Surgeons who love what they do and are therefore committed to teach, want to support every

individual who has the drive to learn this profession. We can teach knowledge and skills to almost all of them. We cannot teach love or motivation. Every single one who brings those unmeasurable qualities deserves remediation. Our ultimate goal with remediation should be competency at the end of residency training and the ability to communicate this competency within the setting of board certification. We should seek non punitive remediation and avoid stigmatization both during and after residency training. We should structure remediation so that it will not disrupt daily work hour flow during residency, and aim to preserve continued practice after completion of residency training. We also should demonstrate compassionate respect in guiding those residents whom we recognize as unsuited for the specialty to find other career paths in medicine. ■

“We can teach knowledge and skills to almost all of them. We cannot teach love or motivation. Every single one who brings those unmeasurable qualities deserves remediation. Our ultimate goal with remediation should be competency at the end of residency training and the ability to communicate this competency within the setting of board certification.”

Christina G. Rehm, MD, FACS, is a Clinical Associate Professor of Surgery at Oregon Health and Sciences University, Portland, OR; Pamela A. Rowland, PhD, is Director, Office of Professional Development and Research Associate Professor in the Departments of Community & Family Medicine and Surgery at Dartmouth Medical School, Hanover, NH.

¹ The American Board of Surgery Booklet of Information 2005: http://home.absurgery.org/xfer/ABS_BookletOfInfo05.pdf (Accessed 3/9/06)

² Rehm CG, Rowland, PA. The remedial year in the general surgery board certification process: How effective is it? *Curr Surg* 2005; 62(6): 644-9.

³ Debas HT, Bass BL, Brennan MF et al: American Surgical Association Blue Ribbon Committee Report on Surgical Education. *Ann Surg* 2004; 241(1): 1-8.

RRC/IRC Column

ACGME approves revisions to program requirements in several specialties and approves two new subspecialties

At its February 2006 meeting, the ACGME approved revisions to the program requirements for Anesthesiology, to be effective July 1, 2008. The Council also approved revisions to the program requirements for the Transitional Year, and revisions to the requirements for the subspecialties of Pediatrics, both with an effective date of July 1, 2007. Revisions to the program requirements for Vascular Surgery also were approved, to become effective July 2006.

The ACGME approved the program requirements for Multidisciplinary Pain Medicine, to become effective July 1, 2007. At that time, the new multi-disciplinary program requirements will replace the pain medicine subspecialty requirements that currently exist in several specialties, such as Anesthesiology, Neurology and Physical Medicine and Rehabilitation.

The Council approved the program requirements for two new subspecialties: Adult Cardiothoracic Anesthesiology, a subspecialty of Anesthesiology, and Congenital Cardiac Surgery, a subspecialty of Thoracic Surgery. Both new sets of requirements became effective February 14, 2006.

The Committee on Program Requirements reviewed the

“The Committee on Program Requirements reviewed the existing ACGME policy on setting the effective dates of new and revised program requirements. It agreed to continue the policy of allowing at least one year for implementation of revisions to existing requirements, but to consider exceptions on a case by case basis.”

existing ACGME policy on setting the effective dates of new and revised program requirements. It agreed to continue the policy of allowing at least one year for implementation of revisions to existing requirements, but to consider exceptions on a case by case basis.

Other news from the February 2006 ACGME meeting

Appointments to Institutional Review Committee

The Board of Directors approved the appointment to the Institutional Review Committee of Dr. John C. Russell, MD, Associate Dean for Graduate Medical Education and Designated Institutional Official (DIO), University of New Mexico, and the reappointment of Carl J. Getto, MD, Senior Vice President for Medical Affairs and DIO, University of Wisconsin Hospitals and Clinics. Both terms are three years from July 1, 2006 to June 30, 2009.

ACGME approves changes to policy and procedure manual

The ACGME Board of Directors approved for review and comment policies addressing dependent subspecialties, a new ACGME Policy to Address Disasters; and the procedures for appeal of expedited adverse actions. The policies will be circulated to all interested parties for review and comment prior to Board approval in June 2006. The Council also approved several corrections and editorial changes in the Bylaws, Policies and Procedures Manual.

ACGME approves four strategic priorities

The Board approved four strategic priorities in the 2006–2008 ACGME Strategic Plan:

- 1) Foster innovation and improvement in the learning environment;
- 2) Increase the accreditation emphasis on educational outcomes;
- 3) Increase efficiency and reduce burden in accreditation; and
- 4) Improve communication and collaboration with key internal and external stakeholders.

The Board of Directors also approved six Enabling Approaches outlined in the Strategic Plan. They include, among others, enhancing support for the ACGME’s review committees, increasing coherence across review committees, aligning the internal structure of the ACGME with the goal of exemplary accreditation; and deploying staff development programs to enable change.

The enabling approaches will allow the ACGME to meet the needs of its internal operations.

Committee on innovation in the learning environment

The Committee on Innovation in the Learning Environment (CILE) reported on its activities, which included updates from two work groups, and a presentation on the activities of the AMA Initiative to Transform Medical Education. One of the workgroups of CILE conducted a focus group exercise with the Council of Review Committee Residents to explore the attributes of the residency experience. This is part of a larger effort to use focus groups to explore the learning environment from the residents’ perspective. The Committee is planning a design conference on the learning environment, to be held September 8–10, 2006, in Rosemont, IL.

Other ACGME Councils and Committees that met at the February Meeting of the Board of Directors included the Monitoring Committee and the Council of Review Committee Chairs. Several committees heard a presentation on the potential of a resident learning portfolio, given by Carol Carraccio, MD, Residency Program Director for Pediatrics at the University of Maryland. Dr. Carraccio developed an extensive learning portfolio for the pediatrics residency program, and is part of the work group developing an ACGME Learning Portfolio (ALP). ■

RRC Development Course Aims to Enhance RRC Preparation for Review of the General Competencies

Jeanne Heard, MD, PhD

The RRC Development Course is proceeding on schedule with the second exercise having been held as part of the ACGME Orientation Workshop for newly appointed Review Committee members on February 11, 2006. More than forty new RRC members participated. The course is designed to assist Review Committee members as the RRCs prepare to comply with the ACGME directive to begin in 2006 citing programs with consequence and to make constructive suggestions.

Carol Carraccio, MD, Director of the Pediatrics Residency Program at the University of Maryland, and Paul Friedmann, MD, Special Advisor to the ACGME, presented an introductory session that provided the context for the groups' discussions. The participants were assigned by specialty category to the following groups: 1) primary care and related specialties, 2) surgical specialties, and 3) hospital based and other specialties. Drs. Carraccio and Friedmann and Rita M. Patel, MD, Assistant Dean for Graduate Medical Education and DIO, University of Pittsburgh Medical Center Medical Education Program, chaired the three groups with assistance from the Executive Directors of the Review Committees.

An additional purpose of these exercises is to demonstrate the need for an integration of the competency requirements with the program requirements and for the related revision of the Program Information Forms (PIFs) to elicit better information from programs regarding the degree to which they have implemented the competencies into their curriculum. ACGME will be partnering with the RRCs as they undertake these revisions in order to provide assistance and to facilitate a greater degree of standardization of requirements and PIFs across the specialties.

The next phase of the course will occur at a meeting of each review committee during 2006 and the first quarter of 2007. One of the seven Development Course Review Group Leaders, who have been trained for this purpose, will conduct the exercise with the assistance of the RRC Executive Director and the RRC members who have participated in one of the two exercises described above. ■

The July Phenomenon: Fact or Fiction

Steven H. Borenstein, MD, Jacob C. Langer, MD

It has been suggested that the extensive turnover of house staff that occurs every July in the numerous teaching hospitals across North America may have a negative impact on patient care. Newly graduated medical students become junior residents, while their more senior colleagues are promoted within the various residency programs. There is a perception among many observers that these eager, bright, but inexperienced doctors deliver suboptimal care during the initial few weeks of the new academic year, and this has come to be known as the "July phenomenon".¹⁻⁴

Table 1
Patient Demographics

	June		July	
	N	%	N	%
Patients				
Neonates	15	18	18	20
Children	68	82	74	80
Total	83		92	
Operations	61		53	
Patient days				
ICU	254	39	302	39
Ward	389	61	474	61
Total	643		776	
High clinical acuity	54		72	
Low clinical acuity	30		36	

To determine whether resident inexperience in July was associated with a reduction in the quality of patient care in our pediatric general surgery inpatient service, we examined the incidence of errors and adverse outcomes during the final month of an academic year (June 2002) and the first month of the subsequent academic year (July 2002).⁵ By evaluating these outcomes, we hoped to determine whether additional mechanisms would be needed to ensure that a uniform, high quality of care is delivered during the preliminary phases of surgical education.

We defined an error as incorrect medical care, whether action or inaction, that had the potential to cause substantial harm. Errors were classified into one of three categories: execution, identification and communication. Execution errors involved an error in carrying out a defined task involving any of the following aspects of patient management: pre-operative investigation/preparation, diagnostic imaging, attaining surgical consent, use of medications, use of operative equipment or personnel and intra-operative decision making. Identification errors involved any error in which a patient's identity was

Table 2
Types of Errors and Error Rate

Errors subtypes	June		July		P Value*
	N	%	N	%	
Execution	31	67	48	83	
Communication	11	24	9	15	
Identification	4	9	1	2	
Total errors	46		58		
Patient days	643		776		
Error rate		7.2		7.5	0.9

*Chi square with continuity correction.

incorrect. Communication errors included failures to correctly convey information between individuals. An adverse outcome was defined as unintended harm to a patient resulting from medical treatment or as a result of the natural history of the disease. The occurrence of an error was not in itself considered an adverse outcome, and not all adverse outcomes were considered to be the result of an error.

The demographics of the patients cared for during the study period are displayed in **Table 1**. Differences in clinical workload during these two months could have affected the numbers of errors and adverse outcomes independent of physician experience. Two parameters were used to compare monthly clinical workload: total patient days in hospital and clinical acuity of the individual patient. Patient days were divided into days in the neonatal or pediatric intensive care unit (ICU) and days on the surgical ward. As a further index of clinical workload, each patient was designated as either high or low clinical acuity based on the admission diagnosis and complexity of inpatient care. The clinical volume in July exceeded that of June by 20% while the clinical acuity of patients seen during the two months was similar.

“During both months, execution errors were the most common, followed by communication errors and then identification errors.”

There were 46 errors in June and 58 errors in July (**Table 2**). During both months, execution errors were the most common, followed by communication errors and then identification errors. To correct for differences in patient volume, the incidence of errors was expressed as a percentage of total patient days. The error rates during the two study months were similar (7.2% in June and 7.5% in July; $P = 0.9$; $RR = 1.05$; $95\% \text{ CI} = 0.72\text{--}1.52$). In June, 24 of 46 errors (52%) were attributed to residents. In July 23, of 58 errors (40%) were attributed to residents. The resident-specific error rates of 3.7% and 3.0% for June and July respectively, were not significantly different ($P = 0.5$; $RR = 0.81$; $95\% \text{ CI} = 0.45\text{--}1.38$) suggesting that a “July phenomenon” did not exist on the pediatric general surgery service at our hospital.

Table 3
Adverse Outcomes

Adverse Outcome	June		July		P Value*
	N	%	N	%	
Unanticipated pain	6		6		
Unanticipated emotional pain	1		2		
Intra-operative problem	0		5		
Unplanned procedure	1		7		
Post operative organ dysfunction	3		14		
Infection	14		9		
Prolonged hospital stay	2		3		
Readmission to hospital	4		6		
Other	1		0		
Total adverse outcomes	32		52		
Adverse outcomes rate	5.0%		6.7%		0.21

*Chi square with continuity correction

There were 32 adverse outcomes in June and 52 adverse outcomes in July (**Table 3**). Most of the adverse outcomes were relatively minor and there was no permanent disability or death. The adverse outcome rates of 5.0% in June and 6.7% in July were not significantly different ($P = 0.21$; $RR = 0.74$; $95\% \text{ CI} = 0.48\text{--}1.13$). Eleven percent (5/46) of errors in June and 22% (13/45) of errors in July resulted in an adverse outcome. While this suggests that July errors were more likely to result in an adverse outcome than June errors, the error-specific adverse outcome rates were not significantly different ($P = 0.21$; $RR = 2.06$; $95\% \text{ CI} = 0.79\text{--}5.36$).

We used the occurrence of errors and adverse events as a measurable outcome to compare the quality of care delivered on a pediatric general surgery service during the final month of an academic year and the first month of the subsequent academic year. If the July phenomenon was real, and the quality of patient care was influenced by physician experience, we would have expected higher rates of errors and adverse outcomes in July. We found that the total error rate, the resident-specific error rate and the adverse outcomes rate during the two study months were similar despite a 20% increase in the clinical volume in July. Our data, therefore, does not support the existence of a July phenomenon and are in keeping with results reported by other groups.^{3,4}

One of the strengths of our study is its prospective design. Retrospective analyses tend to underestimate the true incidence of medical error because the adverse outcome is identified first and then traced back to an error. This process will miss errors that did not result in adverse outcomes (so-called “near-misses”). Interestingly, we found that 80–90% of errors did not result in an adverse outcome suggesting that most errors are either of no clinical consequence or they are corrected by hospital-specific systems before affecting patient care. Obviously, efforts should be focused on further reducing the 10–20% of errors that do lead to adverse outcomes.

There are several explanations as to why we did not find more errors and adverse outcomes in July. Closer supervision by the more senior residents and attending surgeons may have compensated for the relative inexperience of the rest of the surgical team. It is also possible that the residents themselves may have been more cautious in their actions and may have asked for help because of self-awareness of their inexperience. Another factor that may have contributed to our findings is that the residents knew that a study tracking medical errors was being conducted. This is known as the Hawthorne effect and may have caused them to act more carefully and commit fewer errors than they would have otherwise.⁶

“Closer supervision by the more senior residents and attending surgeons may have compensated for the relative inexperience of the rest of the surgical team.”

By defining our outcome measures, we were able to prospectively evaluate the incidence of errors and adverse outcomes on our inpatient service, which served as an index of resident performance and quality of care. Our data indicate that uniform, high quality care was provided regardless of resident experience, which reflects well on the residents who care for the patients and the medical staff that supervises them. ■

Steven Borenstein, MD, was a fellow in the pediatric surgery program at the University of Toronto at the time of the study and is currently Assistant Professor of Surgery at McMaster University and Staff Pediatric Surgeon at the Children's Hospital at McMaster University; Jacob Langer, MD, is Professor of Surgery at the University of Toronto, head of the Division of Pediatric General Surgery at the Hospital for Sick Children, and Program Director for the Pediatric General Surgery Program.

¹ Buchwald D, Komaroff AL, Cook EF, Epstein AM. Indirect costs for medical education. Is there a July phenomenon? Arch. Intern. Med. 1989; 149: 765.

² Rich EC, Hillson S, Dowd B, Gifford G, Luxenberg M. It is OK to get sick in July. JAMA 1991; 265: 212.

³ Claridge JA, Schulman AM, Sawyer RG, Ghezel-Ayagh A, Young JS. The “July Phenomenon” and the care of the severely injured patient: fact or fiction? Surgery 2001; 130: 346.

⁴ Barry WA, and Rosenthal GE. Is there a July phenomenon? The effect of July admission on intensive care mortality and length of stay in teaching hospitals. J. Gen. Intern. Med. 2003; 18: 639.

⁵ Borenstein SH, Choi M, and Langer JC. Errors and adverse outcomes on a surgical service: what is the role of residents? J. Surg. Res. 2004; 122: 162

⁶ Parson H. What happened at Hawthorne? Science 1974; 193: 922.

Linking Educational Processes to Outcome of Care: A Study of Resident Supervision in the Operating Room

Kamal M.F. Itani, MD, FACS, Shukri F. Khuri, MD, FACS

At a time when patient outcome is a measure of performance and pay for performance might become the norm, teaching medical institutions find themselves with a significant dilemma. Our educational endeavor is at the core of our mission, but it might come at the price of worse patient outcomes and less efficient care. At the same time, residency programs are accountable to the ACGME for meeting program requirements designed to assure that residents achieve specified performance outcomes. For example, the ACGME requires that residents should assume a graduated level of responsibility for patients under supervision of physician faculty. By assuming progressively greater

“...the ACGME requires that residents should assume a graduated level of responsibility for patients under supervision of physician faculty. By assuming progressively greater responsibility (under supervision) residents learn how to practice their specialty and under what circumstances to seek assistance from colleagues.”

responsibility (under supervision) residents learn how to practice their specialty and under what circumstances to seek assistance from colleagues. A growing level of independence is based on the level of training and competence as judged by the attending physician. Can both expectations for patient outcomes and resident performance outcomes be achieved or must one come at the expense of the other? How does the ACGME requirement for graduated level of resident responsibility in patient care impact patient outcomes?

While others have investigated this question, our study was unique in several respects:

- 1) The data was adjusted for patient's pre-operative and intra-operative risk factors.
- 2) The data was prospectively collected in a standardized fashion by dedicated nurses, over four years, from multiple hospitals within the same system.
- 3) The sample size of 610, 660 operative procedures provided enough power for solid conclusions.

Our ability to address this question was made possible by the existence of a national database of surgical outcomes, the VA National Surgical Quality Improvement Program (NSQIP). The NSQIP tracks the level of attending supervision of surgical residents in the operating room, patient demographic data, 33 co-morbid conditions, 14 pre-operative laboratory tests and 4 operative and intra-operative variables. Patient outcome variables include 30-day mortality and morbidity, length of hospital stay, and long term survival. The American College of Surgeons has a similar program open to all hospitals.¹ We evaluated the level of attending supervision (patients who were operated on by residents with attendings available to them as opposed to patients who were operated on when the attending was scrubbed or present in the OR) and adjusted for patients' characteristics, comorbidities, pre-operative laboratory values and intra-operative variables.²

The study involved 99 VA hospitals with surgical residency programs affiliated with major academic medical centers that provide clinical education for residents in various surgical specialties. The analysis consisted of a univariate analysis of all measured factors, including pre-operative conditions, comorbidities, and laboratory results as well as intra-operative variables and post-operative outcomes. This was followed by a logistic regression analysis for mortality and morbidity taking into account all factors of significance in the univariate analysis, dividing the patients into two groups: patients who had surgery by residents with attendings available to them and patients who had surgery with attendings present or scrubbed in the operating room. The result of the multivariate regression analysis for 30-day mortality and morbidity are presented

“In the morbidity analysis, the odds of developing a complication were actually lower when the attending was not present in the OR.”

in **Table 1**. In the mortality analysis the odds of a patient dying when the attending was not present in the OR were not significantly different from the odds of dying when compared to the situation when the attending was present in the OR with the exception of the year 2000. In the morbidity analysis, the odds of developing a complication were actually lower when the attending was not present in the OR. These data clearly indicate that the presence or absence of the attending surgeon in the OR did not influence patient outcomes.

Several interesting observations were made from this study:

1. The complexity of cases performed when the attending was not present in the OR was lower. These cases were also performed by more senior residents. This observation reflects the attending surgeon's judgment in attempting to provide more independence to senior residents particularly for less complex cases. It also suggests that the attending surgeons have selected these cases appropriately for the residents to perform, without compromising the outcomes of the patients.

Table 1
Adjusted 30-day mortality and morbidity*

Attending Category	30-Day Mortality** 95% PL			30-Day Morbidity** 95% PL		
	Odds Ratio	Confidence Interval	C-Index	Odds Ratio	Confidence Interval	C-Index
	Level 3 vs. All Other Levels					
1998	0.90	(0.770,1.054)	0.90	0.71	(0.650,0.781)	0.75
1999	0.90	(0.761,1.056)	0.90	0.66	(0.595,0.720)	0.75
2000	0.72	(0.594,0.858)	0.89	0.74	(0.672,0.812)	0.75
2001	0.86	(0.707,1.037)	0.89	0.74	(0.664,0.814)	0.75
2002	1.03	(0.842,1.256)	0.89	0.72	(0.645,0.805)	0.75
2003	0.85	(0.661,1.077)	0.90	0.86	(0.761,0.967)	0.77
2004	0.99	(0.724,1.325)	0.91	1.01	(0.863,1.164)	0.78

* Table from Itani, et al². For additional data, please see Itani, et al².

** Logistic multi-step regression analysis with all predictors of mortality and morbidity, level 3 and all other levels of supervision tested against mortality and morbidity

Reprinted from The American Journal of Surgery, Vol 190, Itani KMF, et al, Surgical resident supervision in the operating room and outcomes of care in Veterans Affairs hospitals, Pages 725-731, Copyright (2005), with permission from Excerpta Medica, Inc."

-
2. The trend of attendings not being present in the OR decreased consistently from 1998 (8.72%) to 2004 (2.69%). Although an increasing level of attending involvement over the years may be explained in part by decreasing work hours after the year 2002, expectations for more accountability by attending physicians to the system have changed substantially in both VA and non-VA teaching sites in recent years.

This study demonstrates a successful approach to evaluating the effect of a general educational process (supervision in the OR) on patient outcomes after surgery, with implications for program quality and program improvement. Other specific questions such as the effect of postgraduate level of training³ or effects of residents' work hours on patients' outcomes⁴ have been recently addressed in studies of large cohorts of patients. As with other fields in medicine, scientific evidence rather than tradition and external pressure should guide our educational processes to provide a safe environment for the patient, the best quality of care and the best trained physicians. ■

Kamal Itani, MD, is Chief of Surgery at the VA Boston Healthcare System and Professor of Surgery at Boston University School of Medicine, and Shukri Khuri, MD, is Professor of Surgery at Harvard Medical School, Boston, MA

¹ American College of Surgeons National Surgical Quality Improvement Program: <https://acsnqip.org/login/default.aspx> (accessed 2/28/06).

² Itani KMF, DePalma RG, Schiffner T, Sanders KM, Chang BK, Henderson WG, Khuri SF. Surgical resident supervision in the operating room and outcomes of care in VA hospitals. *American Journal of Surgery* 2005; 190(5): 725-731

³ Wilkiemeyer M, Pappas TN, Giabbie-Hurder A, Itani KMF, Jannasson O, Neumayer L. Does resident level of training influence the outcomes of inguinal hernia repair? *Annals Of Surgery* 2005; 241(6): 879-884

⁴ Kaafarani MA, Itani KMF, Petersen LA, Thornby J, Berger DH. The impact of resident work hour reduction on surgical outcome. *Journal of Surgical Research* 2005; 126(2): 167-171

Teaching about Transparency: Linking Resident Performance and Patient Care Outcomes

Tina Foster, MD, MPH, MS, Karen E. George, MD, Michele R. Lauria, MD, MS

Good patient care and good learning go hand in hand. For many of us, both patient care and teaching seem to happen episodically. We work in systems that predispose us to see each patient encounter as an isolated event; we thus evaluate our care based on our perceptions of a stream of individual encounters. Our understanding of the small populations we care for is often limited. Similarly, a great deal of teaching occurs during those individual encounters, and many times there is little time for reflection about the systems where residents learn and practice or commonalities among their patients.

“Frequently many different providers are involved in the care of a single patient. Even in cases when we do have data available about a group of patients, how to best use that data may not be clear.”

“Good learning” for residents includes the ability to analyze their own practices and improve their patient care outcomes using a systematic methodology. This can seem a daunting challenge for both residents and faculty. There are many barriers to learning about our own patients. We may not have an assigned panel of patients or group that we feel “belongs” to us. Data may not be readily available to us about demographics, diagnoses, or outcomes. Data may be missing or of poor quality. Working in a supervised environment, residents are not ultimately responsible for decisions about patient care. Frequently many different providers are involved in the care of a single patient. Even in cases when we do have data available about a group of patients, how to best use that data may not be clear. We are often unfamiliar with performance at a departmental or institutional level as well.

We have worked to create a conscious connection for residents between the care that they deliver as individuals, the care we provide as a department, and the care provided by Dartmouth-Hitchcock Medical Center (DHMC). At the institutional level, DHMC recently committed to a “transparency initiative” that includes the posting of outcomes for specific programs on its website. These include both measures which are publicly reported elsewhere, such as those requested by JCAHO or CMS, and measures which have been developed locally because they are important to patients and families. As obstetricians, we are familiar with the great interest

in and public scrutiny of some of our outcomes, such as c-section rates, and thus we were ready to be among the first programs that participated. At the DHMC website, under “Quality Reports”, anyone can view information about pregnancy care at DHMC.¹ Measures include c-section rates, VBAC success rates, epidural rates, patient satisfaction (both inpatient and outpatient) and a variety of others. Although this initiative has received significant attention in the lay press, many residents at DHMC are unaware of it. Our department’s participation has provided an opportunity to begin to explore the implications of such transparency and our roles as members of a community of physicians, midwives, and others.

At a departmental level, we regularly review obstetrical outcomes in our M&M conferences using OBNet, an obstetrical database provided through the Northern New England Perinatal Quality Improvement Network, to record information about all deliveries.² OBNet provides the information needed to compile our monthly statistics and

“Senior residents who are responsible for M&M sessions become familiar with the use of control charts to display important statistics such as c-section rates, induction rates, and operative delivery rates.”

informs both M&M conferences and QA meetings. Senior residents who are responsible for M&M sessions become familiar with the use of control charts to display important statistics such as c-section rates, induction rates, and operative delivery rates. In their discussions of these statistics as well as individual cases, they use concepts of systems based practice to understand both departmental trends and unanticipated outcomes. Although our discussions tend to focus on individual cases rather than aggregated outcomes, we have begun to have productive discussions about departmental trends as well.

Antepartum care by physicians for low-risk pregnancies is currently provided at DHMC by obstetrical teams, each of which includes one or two attending physicians and four residents, one from each year. Residents generally feel a real sense of “ownership” for team patients, which is reinforced by weekly team meetings and chart review. However, most of us do not know the specifics of our team composition. When OBNet became available, we realized this presented an opportunity to share outcomes for each team.

Although most of the outcomes reported by OBNet relate to inpatient care, there is some information directly related to the outpatient settings where the teams function. One metric available is whether Group B Streptococcus (GBS) screening was completed before 37 weeks, in accordance with CDC guidelines. In early 2005 we knew that data on screening for GBS at term would be added to the DHMC public reporting website. We saw this as an opportunity for both residents and attendings to connect their own practices to their team’s performance and to see their results reported publicly in aggregate form. One resident from each team was provided with his or her team’s OBNet report. This resident was asked to review the team’s performance on GBS testing and to work to improve it. Although screening levels were already high, there has been some improvement for the physician teams which include residents. Residents are beginning to see OBNet as a resource that can help them learn more about their practice, rather than a data-entry headache.

This is an early step in connecting resident performance outcomes to improved patient care. Making visible the connection between individual practice and performance at team, departmental, and institutional levels is an essential first step in improving overall outcomes. Ultimately, we want to be able to show that care which involves residents produces superior patient outcomes. Doing this will require thoughtful construction of data sources and providing residents with opportunities to see and evaluate their own results and the skills to work collaboratively to improve them. ■

Tina Foster, MD, MPH, MS, is Associate Director of Graduate Medical Education, Assistant Professor of Obstetrics and Gynecology and Community and Family Medicine, and Associate Director of the Leadership Preventive Medicine Residency Program; Karen E. George, MD, is Assistant Professor of Obstetrics and Gynecology and Program Director, Ob-Gynecology Residency; Michele R. Lauria MD, MS, is Associate Professor of Obstetrics and Gynecology; all are at Dartmouth-Hitchcock Medical Center, Lebanon, NH

¹ Dartmouth-Hitchcock Medical Center home page: <http://www.dhmc.org> (Accessed 3/8/06).

² Northern New England Perinatal Quality Improvement Network: <http://www.nnepqin.org/nneob/servlet/HomePage> (Accessed 3/8/06).

Acknowledgments: In addition to recognizing Pamela Derstine, PhD, for her efforts as Guest Editor, the Editor would like to acknowledge the contributions of Jim Cichon, Amy Dunlap and Tom Mannis in proofing and editing the text, and Nancy VandenBerg and Kurt Wisthuff for layout of the final document.



**Accreditation Council for
Graduate Medical Education**

515 North State Street
Suite 2000
Chicago, Illinois 60610

NON-PROFIT ORG.
U.S. POSTAGE
PAID
ACGME
PERMIT NO. 9451
CHICAGO, IL

Special Issue:
**Improvement Through the
Application of the Competencies**

Change of Address:

Requests for a change of address should be sent to the editor at the address given above. Please include a copy of the mailing label from the most recent copy of the ACGME Bulletin along with your new address.