Assessment of competency in traditional graduate medical education has been based on observation of clinical care and classroom teaching. In anesthesiology, this has been relatively easy because of the high volume of care provided by residents under the direct observation of faculty in the operating room. With the movement to create accountability for graduate medical education, there is pressure to move toward assessment of competency. The Outcome Project of the Accreditation Council for Graduate Medical Education has mandated that residency programs teach six core competencies, create reliable tools to assess learning of the competencies, and use the data for program improvement. General approaches to assessment and how these approaches fit into the context of anesthesiology are highly relevant for academic physicians.

ASSESSMENT of competency in traditional graduate medical education (GME) has been based on observation of clinical care and tests that measure the effectiveness of didactic teaching. In anesthesiology, direct observation of resident performance by staff is the norm, and assessment of competence is often based on global impressions (“I know it when I see it”). In this context, the curricula of anesthesiology residencies have been based on diversity of cases, global assessment, didactic teaching, and measurement of medical knowledge via standard tests such as the in-training examination (ITE), American Board of Anesthesiology written examination or written examinations produced by the Inter Hospital Study Group for Anesthesia Education (Anesthesia Knowledge Tests), or the programs themselves. The penultimate evaluation for many programs has been the 6-month Clinical Competence Committee forms submitted to the American Board of Anesthesiology, although the criteria for “satisfactory” performance are unique to each training program.

Many different forces have created pressure to connect GME with outcomes, \(^1\) including the Federal Government because of their huge financial investment in GME and major industrial organizations with outcome measures as a condition of participation in healthcare contracts. The national interest in patient safety is also linked to measurement of competency.

Rather than wait for a legislative mandate, the Accreditation Council for Graduate Medical Education (ACGME) decided to initiate linkage of GME to outcomes. A comprehensive review of GME was undertaken with the intent to define specific competencies that could be applied to all residents. The result was published in February 1999 as the ACGME Outcome Project. The general competencies are

- patient care
- medical knowledge
- practice-based learning and improvement
- interpersonal and communication skills
- professionalism
- systems-based practice

Full-text definitions for these competencies were published in September 1999, followed by a 10-yr, three-phase timeline for implementation.\(^\dagger\) Any program reviewed after July 1, 2003, was obligated to demonstrate curriculum and assessment of these competencies. The recognition that the response needed to be unique to each medical specialty allows for specific anesthesiology responses within the terms of the ACGME Outcome Project.

The Initial Response within Anesthesiology

Anesthesiology has responded receptively to the Outcome Project. Once the language was published in February 1999\(^\dagger\) it was presented at the 1999 Society for Academic Anesthesiology Chairs/Association of Anesthesiology Programs Directors Meeting as a future direction for the specialty. At the spring 2000 meeting of the Society for Education in Anesthesiology, the Residency Curriculum Committee began to work on a template for compliance. After 4 yr of collaborative work, a preliminary version was made available via the Society for Education in Anesthesiology Web site. This work has evolved into a standing committee as an invited liaison for the specialty to the ACGME.

It may be that the nature of training in anesthesiology...
lends itself to easy acceptance, because a high percentage of direct patient care performed by anesthesiology residents occurs under visual supervision of teaching staff. The blend of clinical care with cognitive and technical teaching is inevitable, because anesthesiology is so intimately tied to acute care medicine. Evaluation of anesthesiology residents by their staff has routinely been based on direct observation of clinical care. The need to transition from global impressions to specific, reliable competency measurement is the challenge for the anesthesiology response to the Outcome Project.

**Principles of Assessment**

The distinction between evaluation and assessment is a movement toward the use of reliable, quantitative tools with a measurable level of objectivity. Assessment can be performed with two different approaches: formative assessment and summative assessment. **Formative assessment** involves collection of information about a student designed to provide feedback and stimulate learning. An example is the review of case totals at the midpoint of a rotation, with the goal of identifying the learning achieved and to influence clinical assignments for the balance of the rotation.

**Summative assessment** is used to make outcome decisions. Because it can be used for adverse actions, standardized written examinations have been a major component because of the need for due process. The downside to using summative assessment in this manner is that the assessment tool drives learning, and students “study to the test” with minimal retention of memorized facts. Perhaps more ominous, standardized examinations may not measure the characteristics that they are used to measure. For the US Medical Licensing Examinations (I–III), what is tested is well defined and should not be used to measure other elements besides the breadth of medical knowledge. There is general evidence that performance on standard examinations can be used to predict clinical performance. However, when anesthesiology standard examinations were carefully reviewed to acquire evidence that would predict dangerous clinical performance, there was no direct correlation with actual performance measures for the same resident. These examinations do predict performance on other standardized tests and measurement of a general fund of knowledge. Written knowledge examinations correlate well with competence for physicians in practice.

There is evidence that challenges the validity of standardized examinations as a measure of clinical performance. High achievement on standardized examinations requires acquisition of knowledge aimed toward the test content but does not necessarily measure higher cognitive functions (e.g., correlation, problem solving). The converse is also true: Faculty who have direct knowledge of clinical performance of residents do not successfully predict their ITE results.

**Changes in the Approach to Assessment in GME**

Global clinical evaluation and standardized testing represent a typical competency measurement in the traditional model of GME. An evolving alternative is assessment, which includes feedback and reinforcement of learning. Knowledge acquisition and demonstration of competence for a complex task involving this knowledge can be different because measurement of the breadth of knowledge may not reflect the ability to use this knowledge to solve problems.

Conditions that facilitate learning are ideal when the clinical experience occurs proximate to the assessment event. Real-time feedback facilitates learning by creating immediate interest in the subject. Assessment tools that are created in an authentic clinical context are more likely to stimulate learning. Nontraditional assessment methods that stimulate learning include self-assessment, peer review, and portfolio. An additional advantage to a realistic context is reinforcement of behavior by the linkage with a task, with intense reinforcement versus “studying to the test” with limited retention.

**Characteristics of an Optimum Assessment Process**

Assessment tools must achieve acceptable levels of performance for six characteristics to be useful. Reliability is the reproducibility of the results. Two different raters should be able to independently measure performance and achieve similar conclusions. Written examinations based on multiple-choice questions (MCQs) are highly reliable in measurement of medical knowledge. Global evaluations of a rotation have low reliability, although this can be improved with intense faculty development to define elements of performance and by having multiple assessments by different raters.

Validity of an assessment tool is determined by whether it actually measures what it is designed to measure. MCQ examinations are thought to have limited validity in predicting clinical competence. Predicting competence in a clinical setting becomes more valid when the assessment occurs in a clinical setting. Global assessments of clinical performance after a rotation have limited validity. Flexibility is determined by how well a given tool can be used to measure performance in different settings. Global evaluation fairs well because of applicability to a wide variety of GME situations. The MCQ examination...
has limited flexibility, reliably measuring medical knowledge but poorly adaptable to other competencies.

Comprehensiveness is related to the extent that an assessment tool measures all elements of performance. Global assessment achieves an acceptable level for comprehensiveness. The use of MCQ examination as a single tool has limited comprehensiveness.

Feasibility is related to whether an assessment tool can be used in any given GME program. Attempts to achieve assessment with high reliability and/or validity have led to the use of standardized patients (SPs) and objective structured clinical examinations (OSCEs). The administrative structure is easy to create when used in large training programs or in medical schools. In the average residency training program, the logistical needs are oppressive, and good tools such as SPs and OSCEs have serious feasibility issues within the anesthesiology world.28

Timeliness of an assessment tool is determined by when the assessment intervention is performed in relation to the measured behavior. The ideal is real-time assessment with immediate feedback. The opposite extreme is the evaluation that occurs weeks or months after the clinical event, resulting in reduced validity, loss of any reinforcement of learning, with rater bias becoming more likely.28

Accountability is the ability to defend the efficacy of an assessment tool. This is especially important for tools used to make summative assessment decisions. These decisions can be adverse and must be defensible as fair and transparent, ideally with guidelines for action that are objective.29

One way to improve assessment is to move from the "pass-fail" habit in GME ("good" or "excellent") to descriptive assessment. For undergraduate medical education, Pangaro50 suggested vocabulary for competence, defining skills by whether the student is a "reporter," an "interpreter," a "manager," or an "educator" of these RIME steps. He advocates measuring skills during performance of a task with the student aware of the assessment. The clinical skills assessment in US Medical Licensing Examinations step 2, Clinical Skills, is an example of a high-stakes, performance-based assessment. The Outcome Project uses descriptive competencies. Other data supports that descriptive assessment is effective in detecting deficiencies in medical knowledge,31 professionalism,52 and patient care.35 The idea that sharing data from sequential descriptive assessments can validate the process has also been previously reported54 and shown to demonstrate face validity.35

Assessment of the Core Competencies in the Context of Anesthesiology

The ACGME has mandated that each program must establish the teaching and assessment of these competencies. The challenge for anesthesiology is to make this practical and feasible in the context of anesthesiology residency. The amount of direct supervision and observation of patient care is high within anesthesiology, and the teaching of the competencies should be easily accomplished. Assessment is more problematic. The requirements of the Outcome Project make it mandatory that training programs measure learning and use the data for remediation of individual residents and process improvement of the training program. This requires anesthesiology program directors to select reliable assessment tools for each of the six competencies with a reasonable degree of feasibility. The medical education literature has a large number of reports about assessment tools, and it is worthwhile to define those tools that could potentially be used in anesthesiology residency programs.

Outcome Project—General Competencies: 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. Residents are expected to

- communicate effectively and demonstrate caring and respectful behaviors when interacting with patients and their families
- gather essential and accurate information about their patients
- make informed decisions about diagnostic and therapeutic interventions based on patient information and preferences, up-to-date scientific evidence, and clinical judgment
- develop and carry out patient management plans
- counsel and educate patients and their families
- use information technology to support patient care decisions and patient education
- perform competently all medical and invasive procedures considered essential for the area of practice
- provide healthcare services aimed at preventing health problems or maintaining health
- work with healthcare professionals, including those from other disciplines, to provide patient-focused care

Written Examinations

Knowledge is required for good patient care, but measurement of knowledge alone does not directly evaluate patient care skills. A modification of the written examination has been described, where the test is created from distinctive patient scenarios and the questions are designed to cover unique complaints in a surgical clerkship.36

Multiple-choice questions are relatively common assessment tools used in anesthesiology programs. Realis-
Objective Structured Clinical Examination

The OSCE has been advocated as a means of measuring patient care skills.\(^7\)\(^-\)\(^9\) It can be adapted to the clinical reality of a variety of specialties, including surgery,\(^41\)\(^-\)\(^44\) internal medicine (IM),\(^45\)\(^,\)\(^46\) pediatrics,\(^47\)\(^-\)\(^52\) family medicine (FM),\(^53\) and various clinical settings\(^45\)\(^,\)\(^54\) including both the outpatient\(^42\) and hospital-based practices.\(^55\)\(^,\)\(^56\)

In addition to assessment, OSCE enhances learning because of immediate feedback.\(^57\) The OSCE format has been applied to the direct assessment of patient care in geriatric medicine,\(^58\) emergency medicine (EM),\(^59\) psychiatry,\(^60\)\(^,\)\(^61\) obstetrics and gynecology,\(^62\)\(^,\)\(^63\) and rheumatology.\(^64\)

Technical (e.g., simulation of laparoscopy) and bench elements (e.g., identification of tissue with a microscope) can be added in the OSCE format\(^55\)\(^-\)\(^68\) for surgery as objective structure assessment of technical skills, although surgery residents often do not believe this to be a valid measurement of either knowledge or technical skill.\(^59\) The OSCE does discriminate clinical knowledge and technical skills criteria by level of experience (physician assistant, medical student, surgical resident).\(^70\) Entry-level clinical skills can be measured reliably with OSCE.\(^66\)\(^,\)\(^71\) When SPs are used in the OSCE format, they can be trained to provide feedback about the clinical skill of the student.\(^72\) Candidates accept the OSCE format with a high level of satisfaction, reporting testing as an active learning experience.\(^73\)

There has been limited enthusiasm for OSCE in anesthesiology programs. There are serious issues with cost, feasibility,\(^74\)\(^-\)\(^77\) and reliability.\(^41\)\(^,\)\(^45\)\(^,\)\(^52\)\(^,\)\(^62\)\(^,\)\(^78\) OSCE programs in the primary care settings (IM, FM, pediatrics) are supplemented by SPs, which are easy to recruit in these specialties, but not in anesthesiology. There are a limited number of anesthesiology programs that use OSCE for assessment of patient care, and evidence of the efficacy and reliability of OSCE in this setting remain unpublished at this time.

Direct Observation

Direct observation is probably the most frequently used assessment tool for patient care skills in anesthesiology. The reliability improves when the person performing the assessment is not directly involved in the clinical care, requiring additional resources.\(^79\)\(^,\)\(^80\) The optimum feedback format is written, if the goal is to stimulate learning.\(^81\) Validity of global assessment improves when structured criteria are used.\(^23\) An example of structured criteria is the RIME terminology of Pangaro,\(^2\) which uses descriptive terms for ascending skill levels of performance.\(^2\) Without structure, the majority of strengths/weaknesses were missed by experienced IM faculty compared with assessments with a structured format.\(^82\)\(^,\)\(^83\) The reliability issue of the “easy grader” is magnified with global ratings of observed performance.\(^84\)\(^,\)\(^85\) Interrater reliability is low even with extensive faculty training.\(^86\) Longitudinal observations using a template yield superior results compared with observations without a template.\(^87\) Preestablished criteria for direct observation are valuable to identify skill levels and the need for remediation of highly technical tasks.\(^88\)\(^,\)\(^89\) Direct observation on multiple occasions by the same observer is more reliable than observation on a single occasion.\(^90\) The reliability of multiple encounter observation is better if criteria are rigid when different observers are used.\(^91\)

Direct observation of anesthesiology residents for assessment of patient care was indirectly validated by Rhoton et al.,\(^92\) who observed a correlation between observed deficiencies in noncognitive skills (confidence, composure, eagerness to learn, interpersonal skills, willingness to take instruction, professional behavior) and critical incidents. The potential for the validation of direct observation of patient care as an assessment tool is excellent if it is linked with simulation using the same observers. Direct observation of clinical performance by an observer not involved in the patient care is an option to improve reliability.

Self-assessment

A relatively underexplored area of assessment of patient care is self-assessment. Accurate self-assessment skill does not come naturally and requires training. Residents were able to arrive at the same evaluation of technical skills as their teachers with a modest amount of training,\(^93\) especially if the training included explicit expectations.\(^94\) An added advantage is the additional learning from the act of self-assessment.\(^95\) Specific training for reflection improves the ultimate product in a system of self-assessment.\(^96\) In an obstetrics and gynecology rotation, reflection was taught using the medical literature and applied to clinical situations, improving the student’s ability to evaluate their own performance.\(^97\) In a general practice setting, reflection about challenging cases combined with journaling and third-party feedback improved self-assessment skills.\(^98\) Student performance on self-assessment activities matched their progress in clinical skill acquisition.\(^99\) Oral surgery residents were able to accurately identify areas of skill in which they required more experience and teaching.\(^100\) When initial attempts at self-assessment by residents were compared with subsequent attempts, training and
repetition resulted in improved skill. Self-assessment may be more effective when combined with auditing and feedback for residents. In general, trained self-assessment is harsher than faculty assessment of the same event.

Self-assessment has not made significant inroads within anesthesiology education, although one report of self-reporting of medical errors suggested good educational merit. A monitoring process could also have the same effect. There is potential for self-assessment by anesthesiology residents, if explicit criteria are created by the program along with clear definitions of the evidence that could be used by the resident to establish competency.

**Standardized Patients**

Use of SPs has been widely accepted as a means of assessing patient care in undergraduate medical education. SPs can be adults or children, although children as subjects have feasibility issues. Providing the SP with a simple script makes the interaction more active in assessing consultation skills. In a highly controlled application, SP provided an effective part of an assessment tool for patient care skills of surgical residents. SPs also proved to be a reliable means of assessing technical skills in an EM training program. In an IM residency, global evaluations were compared with assessment of clinical skills using SPs and were found to have low correlation, suggesting that the SP experience was measuring something different. Videotape review of actual patient care in a postresidency setting was found to be more effective as an assessment tool based on feasibility compared with SP stations. In a medical student setting, SP performance correlated well with clinical performance.

There has been limited application of SPs within anesthesiology because of feasibility issues. The number of students is large in the undergraduate setting, justifying the effort and expense to locate and maintain these patients. The faculty-to-student ratio makes the resource expenditure for faculty development reasonable. It is also practical in the IM, psychiatry, and FM residencies because the patients can be easily recruited from continuity clinics, although faculty development effort is considerable. The most relevant clinical situations within anesthesiology to be evaluated do not easily lend themselves to the SP format (except possibly chronic pain management), and recruitment of SP is not well suited to most anesthesiology residency settings.

**Audits**

The combination of examining medical records combined with targeted feedback makes auditing an effective tool for assessment of specific elements of patient care. The completeness of physical examination can be assessed by audit in a primary care setting. In clinical settings, auditing is a highly effective assessment tool with the additional advantage of encouraging the preferred clinical behavior.

Auditing is a regular part of the practice of anesthesiology for administration, billing, and appropriate use of controlled substances. Auditing of anesthesia records for assessment of patient care skills would have a very limited return for the effort, unless combined with a structured tool to measure a specific outcome.

**Simulation**

To simplify the demands of creating an OSCE, or recruiting and maintaining SPs, there has been a sustained effort to create realistic clinical situations electronically to both teach and assess patient care. Human patient simulation has demonstrated considerable efficacy for assessment of technical elements of patient care, such as emergency thoracotomy, bronchoscopy, endoscopy, laparoscopy, lumbar puncture, and various surgical maneuvers. It has also proven effective for measurement of rapid problem-solving skills in the acute care context.

Simulation has demonstrated considerable promise in anesthesiology for the teaching and assessment of the management of acute clinical crisis, similar to the simulation of aviation “near-misses.” A logical extension would be the use of simulation for certification, which has been implemented for medical licensure in Italy. In a surgical residency, simulation performance correlated well with global clinical evaluations of technical performance in the operating room. In a medical student setting, simulation assessment was compared with global evaluation and SP performance and found to correlate well. Assessment using simulation enhances learning in a way not achieved by didactic teaching and textbooks. Simulation may be an ideal approach to the measurement of acute care skills in anesthesiology. Defining behavior related to critical incidents may provide a unique means of assessment of anesthesiology residents.

**Outcome Project—General Competencies: Medical Knowledge**

Residents must demonstrate knowledge about established and evolving biomedical, clinical, and cognate (e.g., epidemiologic and social–behavioral) sciences and the application of this knowledge to patient care. Residents are expected to

- demonstrate an investigatory and analytic thinking approach to clinical situations
- know and apply the basic and clinically supportive sciences which are appropriate to their discipline

**Multiple-choice Question Examinations**

The role of the standard written examination using well-constructed MCQs remains the accepted standard.
for measuring breadth of knowledge. For physicians in practice, a comprehensive written examination compares well to other assessment formats, suggesting MCQs as a practical tool.\textsuperscript{147} Aside from breadth of knowledge, it is not clear what standard examinations measure, or whether they are a reliable means of high-stakes outcome decisions.\textsuperscript{148} True–false questions probably should not be used.\textsuperscript{149} When traditional MCQ tests were compared with OSCE short essay and extended matching for medical students, each measured a different subset of knowledge.\textsuperscript{150} Using written ITEs in conjunction with some other tools may be a better approach to achieve comprehensive assessment of medical knowledge.\textsuperscript{20} In a radiology residency, global performance evaluation was used to predict ITE results, with poor correlation.\textsuperscript{151} ITE performance in a psychiatry residency predicted cognitive skills but did not necessarily predict clinical skills.\textsuperscript{152} ITE scores did not correlate with global performance evaluations in a surgery residency.\textsuperscript{153} An IM residency reviewed\textsuperscript{142} multiple assessment instruments and concluded that combinations of tools were needed to achieve comprehensive assessment.\textsuperscript{45} The lack of correlation between assessment of knowledge and other measures led a surgery residency to advocate the use of multiple tools to achieve comprehensive assessment.\textsuperscript{155}

It is not clear how well measurement of knowledge via MCQs translates to application of medical knowledge.\textsuperscript{154,155} Cox et al. looked at the clinical impression of knowledge from program directors compared with the results of the ITE and found a high level of correlation.\textsuperscript{8} This was also reported in a radiology residency.\textsuperscript{154,156} There was a low correlation with knowledge assessed using SPs in undergraduate medical education compared with written examination results.\textsuperscript{107} Case presentation can be used for evaluation of medical knowledge with high reliability, especially if the elements presented are measured against a template.\textsuperscript{155,156} When MCQs were compared with other styles of assessment (audit, global rating, SP) in an IM residency, it was clear that different elements of training were being assessed and that no one tool was comprehensive.\textsuperscript{122} For physicians in practice, self-assessment using MCQs seems to be an excellent approach to continuous professional development in a rheumatology setting.\textsuperscript{157}

Multiple-choice question tests have a traditional role in the assessment of medical knowledge within anesthesiology. The ITE examinations can be used for formative assessment and remediation based on the key words for incorrect answers. Some anesthesiology training programs measure the progress of acquisition of knowledge using the Anesthesia Knowledge Test at 1, 6, and 18 months of training. A small number of programs generate internal written examinations used for summative assessment. Standard examinations are considered a reliable measure of the breadth of knowledge of anesthesiology residents, although not necessarily the depth of knowledge, which probably should be measured with another tool.

**Oral Examination**

Oral examinations have a role in assessment of medical knowledge that is distinct from standardized written examinations.\textsuperscript{158} Resources must be invested in faculty development to ensure reliability, because oral examinations require both questions and human examiners.\textsuperscript{159} The conduct of the examination as well as the examiners must be structured to ensure standardization between candidates.\textsuperscript{160} If oral examinations are to be used for summative assessment, previous exposure to the format in a lower consequence setting is essential, because previous experience with oral examinations format may be minimal.\textsuperscript{161,162}

There is good evidence that oral examinations can provide a valid measure of some elements of medical knowledge, despite concerns about the reliability.\textsuperscript{163} The variability between oral examinations (reliability) is acceptable,\textsuperscript{164} correlates well with other criteria of medical knowledge within anesthesiology, and functions reliably in the setting of an anesthesiology residency.\textsuperscript{165} The American Board of Surgery oral examination has also been shown to correlate well with other measures of performance.\textsuperscript{166}

**Outcome Project—General Competencies:**

**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. Residents are expected to

- analyze practice experience and perform practice-based improvement activities using a systematic methodology
- locate, appraise, and assimilate evidence from scientific studies related to their patients' health problems
- obtain and use information about their own population of patients and the larger population from which their patients are drawn
- apply knowledge of study designs and statistical methods to the appraisal of clinical studies and other information on diagnostic and therapeutic effectiveness
- use information technology to manage information, access online medical information, and support their own education
- facilitate the learning of students and other healthcare professionals

There is increasing evidence that previous experiences can influence subsequent clinical performance, if the experiences are properly observed and subject to reflec-
tion. Bad behavior in the clinical settings has a clearly negative impact on patient care as well as providing an unprofessional role model to those team members still in training. Even the most inexperienced member of a clinical service can be trained to recognize unethical conduct. Mentorship about the expectations of ethical behavior in the clinical setting creates both learning and assessment of this element of practice-based learning and improvement (PBLI). There have also been reports of assessment of evidence-based medicine programs, including audits in primary care practice, an evidence-based medicine skills test for IM residents, and a Web exercise.

Evidence-based medicine has become a regular element of the practice of anesthesiology with the creation of numerous practice guidelines. A measure of PBLI could be derived from audits of clinical practice where these guidelines apply (e.g., Pre-Anesthesia Testing) although reports of this approach have not been published to date.

Mentorship

One of the oldest forms of PBLI has been mentorship. A trainee has clinical experience, shares it with a senior physician, and receives feedback that leads to improvement. This learning and assessment loop has been described for EM. Use of mentorship in this manner has not been reported for anesthesiology but certainly could be studied, particularly if combined with mentorship of clinical care in the simulation setting.

Self-reporting

One potentially excellent format for assessment of PBLI is self-reporting of elements of patient care. This would accomplish learning and assessment in tandem. Review of critical events is an excellent form of PBLI triggered by self-reporting. This approach has been validated in an anesthesiology residency for self-reporting of medical errors. Self-reporting is especially relevant because physicians consistently identify the review of critical incidents and medical errors as the most significant impetus for change. Self-reporting combined with peer review can be used for the assessment of PBLI for the underperforming physician. Self-reporting with group discussion results in PBLI assessment via comparison feedback.

Videotape and information management technology are excellent adjuncts to the assessment of PBLI by improving the accuracy of self-reporting. Handheld computers also have the potential for recording clinical information, which were used after the completion of patient care to measure PBLI in an FM residency setting. All informatics have the potential to improve assessment of PBLI in GME.

Self-reporting is a regular part of continuous quality improvement programs that exist in virtually every anesthesiology practice. If self-reporting were combined with some other form of active data collection, it could be used by faculty for feedback that would likely be very effective for assessment of PBLI.

Outcome Project—General Competencies: Interpersonal and Communication Skills

Residents must be able to demonstrate interpersonal and communication skills that result in effective information exchange and teaming with patients, their patients families, and professional associates. Residents are expected to:

- create and sustain a therapeutic and ethically sound relationship with patients
- use effective listening skills and elicit and provide information using effective nonverbal, explanatory, questioning, and writing skills
- work effectively with others as a member or leader of a healthcare team or other professional group

Objective Structured Clinical Examination

The OSCE is a reliable means of measuring communication skills. Patient outcomes improve with excellent physician–patient interaction. These same variables can be measured by OSCE assessment or video-assisted OSCE. Effective written communication can be taught and measured in an OSCE format that focuses on written communication skills. OSCE format can be used to measure language skills in international medical school graduates.

Case presentation is a less demanding alternative to OSCE for measuring verbal communication skills. Creating scripts for interacting with patients will improve communication, as long as it is observed. Sessions where residents viewed tapes of difficult patient interviews resulted in better subsequent patient interviews compared with interviews by residents without the teaching intervention.

In an excellent example of an integrated project for teaching and assessment, Morgan and Winter reported a three-step process starting with a formal presentation of expectations, followed by an interactive seminar, and a session that focused on problem solving in a pediatric residency. A similar multistep teaching process significantly improved interview skills compared with a control group in an IM residency.

Objective structured clinical examination has not been reported as a tool for assessment of communication skills in anesthesiology residents. Case presentation, however, is a universal part of anesthesiology residency and subject to global assessment, although the results have not been reported. Perhaps most potentially useful would be...
assess the communication skills of residents in a wide range of settings, including [details about different settings].

Peer Review

Properly structured peer review yields useful assessment information about communication skills for physicians in practice, and for residents in training. The structured peer-review format has limited applicability in anesthesiology training programs, because of the scarcity of situations where residents share the same task, unlike surgery or IM services where groups of residents function as a team. The structured peer review (“snapshot”) is being used in some anesthesiology programs in those areas where there is a high contact level between the residents and those being asked to use the assessment tool, such as preanesthesia testing clinics, postanesthesia care units, intensive care units, and pain management centers.

Outcome Project—General Competencies: Professionalism

Residents must demonstrate a commitment to carrying out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population. Residents are expected to

- demonstrate respect, compassion, and integrity; a 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 ongoing professional development
- demonstrate a commitment to ethical principles pertaining to provision or withholding of clinical care, confidentiality of patient information, informed consent, and business practices
- demonstrate sensitivity and responsiveness to patients’ culture, age, sex, and disabilities

Numerous guidelines and standards for ethics and professional behavior of physicians in general and in a variety of specialties, including EM, orthopedic surgery, obstetrics and gynecology, and in a variety of specialties, including EM, orthopedic surgery, obstetrics and gynecology, have been published. More challenging is using these general resources to measure professional behavior of physicians in practice. Peer assessment and self-assessment, the OSCE format, patient feedback, role models, SPS, and simulation have been used to measure ethical behavior and professionalism.

Case-based problem resolution is also a means to measure professionalism. Active intervention to resolve episodes of unprofessional behavior is also an effective means to assess professionalism.

Comprehensive review in the 360-degree format has been reported for assessment of professionalism in medical students, for physical medicine and rehabilitation residents, and for radiology residents. These 360-degree evaluations yielded data, but it required considerable effort (feasibility), resulting in a limited amount of new information. The opposite extreme has also been reported, with episodes of unprofessional behavior correlating with critical incidents. Identifying residents with behavior issues in clinically relevant settings has been described for EM residents.

Global assessment of professionalism is a required part of regular resident assessment within anesthesiology (Acquired Characteristics). Establishing more substantial data for comprehensive assessment of professionalism is an important goal.

Outcome Project—General Competencies: 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. Residents are expected to

- understand how their patient care and other professional practices affect other healthcare professionals, the healthcare organization, and the larger society and how these elements of the system affect their own practice
- know how types of medical practice and delivery systems differ from one another, including methods of controlling health care costs and allocating resources
- practice cost-effective healthcare and resource allocation that does not compromise quality of care
- advocate for quality patient care and assist patients in dealing with system complexities
- know how to partner with healthcare managers and healthcare providers to assess, coordinate, and improve health care and know how these activities can affect system performance

Systems-based practice is perhaps the most difficult of the competencies for assessment within anesthesiology, because the focus is on being able to interface effectively with healthcare systems. This is a challenge for anesthesiology training programs, because of the limited focus outside the operating room.

One example of the assessment of systems-based practice is peer review. For physicians in practice, the peer review was most effective when combined with written feedback, including review by partners, referring physicians and patients in a primary care practice setting. In another setting, peers were selected by the evaulce, questionnaires were sent by mail, and the information sought was open-ended. The unstructured format pri-
A Comprehensive Portfolio Approach to Competency Assessment

One attractive approach to satisfying the assessment need for the Outcome Project is a comprehensive portfolio assessment tool. The use of portfolios derives from the graphic arts and has been successfully adopted in professional training and assessment in a wide range of fields.228,229 There are several institutions in the United Kingdom and The Netherlands leading medical education in the use of student portfolios.230–234

The starting point for portfolio assessment in medical education is to define performance in terms of competencies, such as the six competencies in the Outcome Project. The next step is to define standards within these competencies and the kind of evidence that can be used to demonstrate mastery of these standards. In an active portfolio system, the student is responsible to select the evidence to demonstrate mastery, often accompanied by written demonstration (essay) or oral defense of performance. In a passive portfolio system, the evidence is assembled in a similar manner for all being assessed. For summative assessment, the portfolios are reviewed by a group of experts. Before examination of any portfolio, the assessment group reviews each standard to establish a common definition of mastery for each. It is then possible to review each portfolio and, for each competency, define whether the individual student has met the standard, not met the standard, or not provided sufficient evidence. For the Outcome Project, this kind of portfolio assessment could be applied to individual or a subset of competencies, or become the primary means of assessing all of the competencies.

Adoption of the portfolio approach has in part been driven by the search for a tool that encourages reflection and that requires active participation by students in the assessment process.235–238 Reflection is a valuable tool within portfolio assessment because it drives the student to use evidence to improve their own performance and learn in the process. Reflection and self-assessment are key concepts in portfolio assessment systems.239–242 The process of determining mastery of each standard is ideally suited to the creation of a learning plan to modify subsequent training for the resident, and when this feedback is assembled cumulatively for a group of residents, it is well suited for use in program improvement.

The portfolio can be used as a tool for assisting with both formative and summative assessment. During formative portfolio review, students reflect on assessment evidence from their course work and feedback from faculty to self-evaluate progress and set learning goals.241 In this process, ensuring that appropriate progress is occurring and setting learning goals that specify activities addressing areas of weakness is essential.254 When portfolios are used for summative assessment, the portfolio review must determine whether the student has achieved the determined level of mastery of competencies, and this in turn dictates promotion decisions.242

The feasibility of portfolio assessment can be problematic because a large amount of data must be assembled for each portfolio and the review process requires considerable faculty effort.245 The technical difficulty of accumulating the data can be improved with computerization.244 Paper-based portfolios are large, and review for assessment is difficult. These feasibility issues in turn create serious validity concerns. Reliability of portfolio assessment has been challenged when the available evidence is limited.245 Some portfolio assessment projects have been reported in GME, including psychiatry,246 and EM.247 Higher test scores as evidence of improved learning as a result of portfolio assessment have been reported in undergraduate medical education.248 The amount of information needed to evaluate a portfolio and the number of faculty to read the portfolio has been reported from a psychiatry residency.249 The use of one

No single measure of anesthesiology practice performance is likely to comprehensively measure physician interaction with the healthcare system as a whole.158 This is especially true with global assessment for this competency, because the traditional equation of “performance equals competence” has been challenged.227

Physician response to community and governmental pressures is an element of systems-based practice that is easy to recognize, harder to define, and problematic for assessment.222–224 Use of continuing medical education as a criterion for licensure is a more tangible means of defining expectations.224 Combining multiple expectations may be the most effective means of ensuring practice change.225 Pressure for quality criteria will likely become a part of continuing medical education, driving continuing medical education providers to focus learning encounters toward these goals and measure subsequent outcomes.226

Continuous quality improvement performance review and incident analysis are fundamental parts of health care and an example of systems-based practice. Changes in practice are inevitable when desired outcomes are defined.221 Use of videotape and information handling technologies can achieve assessment of various endpoints.177 Physician response to community and governmental pressures is an element of systems-based practice that is easy to recognize, harder to define, and problematic for assessment.222–224 Use of continuing medical education as a criterion for licensure is a more tangible means of defining expectations.224 Combining multiple expectations may be the most effective means of ensuring practice change.225 Pressure for quality criteria will likely become a part of continuing medical education, driving continuing medical education providers to focus learning encounters toward these goals and measure subsequent outcomes.226

Peer review has been used for evaluation of the underperforming physician as a means of both assessment and performance improvement.175 When peer review is combined with group feedback, assessment and practice improvement occur.176

The feasibility of portfolio assessment can be problematic because a large amount of data must be assembled for each portfolio and the review process requires considerable faculty effort.245 The technical difficulty of accumulating the data can be improved with computerization.244 Paper-based portfolios are large, and review for assessment is difficult. These feasibility issues in turn create serious validity concerns. Reliability of portfolio assessment has been challenged when the available evidence is limited.245 Some portfolio assessment projects have been reported in GME, including psychiatry,246 and EM.247 Higher test scores as evidence of improved learning as a result of portfolio assessment have been reported in undergraduate medical education.248 The amount of information needed to evaluate a portfolio and the number of faculty to read the portfolio has been reported from a psychiatry residency.249 The use of one
portfolio process to assess all six competencies has been described in a psychiatry residency. The ACGME is sponsoring a portfolio-design project at several sites, with the intention of creating a structure with the flexibility to be implemented at any ACGME-accredited residency to achieve comprehensive assessment.

The portfolio assessment approach to competency assessment has the potential to be highly useful in anesthesiology residencies. The challenge will be defining the competencies and collecting the type of evidence that can be used by the resident to establish competency. It may be that a portfolio of competencies could be combined with a form of active defense analogous to a thesis defense in graduate school education.

Conclusion

The evolution away from global evaluations (“I know it when I see it”) and MCQ examinations (“My score is . . .”) to competency-based assessment is a natural evolution in GME. The transition within anesthesiology should be smooth because of the high volume of direct observations of resident performance and the daily evaluation of medical knowledge, communications skill, and professional behavior that is an inevitable part of acute care medicine. Rose and Burkle suggest that it is apparent that what we have been doing for years (the American Board of Anesthesiology Clinical Competence Reports) maps directly to the Outcome Project, in a manner that may even be complementary. The teaching of the Outcome Project competencies should be straightforward. Assessment of learning and using these data to change individual and program outcomes is more challenging. For each competency, a number of different assessment tools can be applied, with variable kinds of data resulting. Specific circumstances of individual programs must determine which tools are used, and unique applications of these tools may need to be created to fit their clinical setting. Combining resources into a comprehensive portfolio assessment may prove to be the best means to link teaching, learning, assessment, outcome, and systematic process improvement within GME and, specifically, within anesthesiology. This could also provide the linkage suggested by Rose and Burkle between the American Board of Anesthesiology data and the ACGME competencies, further reinforcing the optimum cycle of assessment that encourages learning.

References


Gilliland DS, Schwartz MA. Assessing clinical performance with standardized patients. JAMA 1997; 278:79–91


Fleece TH, Manyon AT, Serwos T, Pannazera KJ. Toward validation of an assessment tool designed to measure medical students’ integration of scientific knowledge and clinical communication skills. Eval Health Prof 2003; 26:222–35


Ramsdell JW, Berry CC. Evaluation of general and traditional internal medicine residents utilizing a medical records audit based on educational objectives. Med Care 1983; 21:1144–53

Gonzalez AM, Cushieri A. Assessment of innate ability and skills for medical licensing examination using computer-based case simulations and standardized patients. Acad Med 2002; 77:87–90


Des Marchais JE, Jean P. Examiners of examiner training on open-ended, higher level taxonomic level questioning in oral examinations. Teach Learn Med 1993; 5:24–8


Assessment in Anesthesiology, V 106 No 4 Apr 2007


