

## Assessment of Competency in Anesthesiology

John E. Tetzlaff, M.D.\*

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,<sup>1</sup> 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.<sup>†</sup> 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,<sup>†</sup> 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

\* Professor of Anesthesiology, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, and Vice Chair for Education, Division of Anesthesiology, Critical Care Medicine and Comprehensive Pain Management, Cleveland Clinic.

Received from the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, and Center for Anesthesiology Education, Division of Anesthesiology, Critical Care Medicine and Comprehensive Pain Management, Cleveland Clinic, Cleveland, Ohio. Submitted for publication May 1, 2006. Accepted for publication November 16, 2006. Support was provided solely from institutional and/or departmental sources.

Address correspondence to Dr. Tetzlaff: Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195. tetzlaj@ccf.org. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

† Available at: www.acgme.org/Outcome. Accessed November 30, 2006.

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.<sup>2</sup> 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<sup>3</sup> and should not be used to measure other elements besides the breadth of medical knowledge.<sup>4</sup> There is general evidence that performance on standard examinations can be used to predict clinical performance.<sup>5</sup> 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.<sup>6</sup> These examinations do predict performance on other standardized tests<sup>7</sup> and measurement of a general fund of knowledge.<sup>8</sup> Written knowledge examinations correlate well with competence for physicians in practice.<sup>9</sup>

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).<sup>10</sup> The converse is also true: Faculty who have direct knowl-

edge of clinical performance of residents do not successfully predict their ITE results.<sup>11</sup>

## 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,<sup>12</sup> which includes feedback and reinforcement of learning.<sup>13</sup> Knowledge acquisition and demonstration of competence for a complex task involving this knowledge can be different<sup>14</sup> because measurement of the breadth of knowledge may not reflect the ability to use this knowledge to solve problems.<sup>15</sup>

Conditions that facilitate learning are ideal when the clinical experience occurs proximate to the assessment event.<sup>15</sup> Real-time feedback facilitates learning by creating immediate interest in the subject.<sup>15</sup> Assessment tools that are created in an authentic clinical context are more likely to stimulate learning.<sup>15</sup> Nontraditional assessment methods that stimulate learning include self-assessment,<sup>16</sup> peer review,<sup>17</sup> and portfolio.<sup>18</sup> An additional advantage to a realistic context is reinforcement of behavior by the linkage with a task,<sup>12</sup> with intense reinforcement<sup>19</sup> versus “studying to the test” with limited retention.<sup>20</sup>

## Characteristics of an Optimum Assessment Process

Assessment tools must achieve acceptable levels of performance for six characteristics to be useful.<sup>20</sup> *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<sup>21,22</sup> and by having multiple assessments by different raters.<sup>23,24</sup>

*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.<sup>4</sup> Predicting competence in a clinical setting becomes more valid when the assessment occurs in a clinical setting.<sup>25–27</sup> Global assessments of clinical performance 7–14 days after a rotation have limited validity.<sup>26</sup>

*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.<sup>28</sup>

*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.<sup>28</sup>

*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.<sup>29</sup>

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, Pangaro<sup>30</sup> 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,<sup>31</sup> professionalism,<sup>32</sup> and patient care.<sup>33</sup> The idea that sharing data from sequential descriptive assessments can validate the process has also been previously reported<sup>34</sup> and shown to demonstrate face validity.<sup>35</sup>

### 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 compe-

tencies. 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.<sup>36</sup>

Multiple-choice questions are relatively common assessment tools used in anesthesiology programs. Realis-

tic clinical scenarios are used for some of the stems of written examination questions on the anesthesiology ITE and the American Board of Anesthesiology Written Board Examination. This will increase as the extended match format is added to the ITE/written board question pool. Although MCQ examinations are attractive to program directors because of their reliability, there is serious doubt about the validity of using MCQs to assess patient care skill.

#### *Objective Structured Clinical Examination*

The OSCE has been advocated as means of measuring patient care skills.<sup>37-40</sup> It can be adapted to the clinical reality of a variety of specialties, including surgery,<sup>41-44</sup> internal medicine (IM),<sup>45,46</sup> pediatrics,<sup>47-52</sup> family medicine (FM),<sup>53</sup> and various clinical settings<sup>45,54</sup> including both the outpatient<sup>42</sup> and hospital-based practices.<sup>55,56</sup> In addition to assessment, OSCE enhances learning because of immediate feedback.<sup>57</sup> The OSCE format has been applied to the direct assessment of patient care in geriatric medicine,<sup>58</sup> emergency medicine (EM),<sup>59</sup> psychiatry,<sup>60,61</sup> obstetrics and gynecology,<sup>62,63</sup> and rheumatology.<sup>64</sup>

Technical (*e.g.*, simulation of laparoscopy) and bench elements (*e.g.*, identification of tissue with a microscope) can be added in the OSCE format<sup>65-68</sup> 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.<sup>69</sup> The OSCE does discriminate clinical knowledge and technical skills criteria by level of experience (physician assistant, medical student, surgical resident).<sup>70</sup> Entry-level clinical skills can be measured reliably with OSCE.<sup>66,71</sup> When SPs are used in the OSCE format, they can be trained to provide feedback about the clinical skill of the student.<sup>72</sup> Candidates accept the OSCE format with a high level of satisfaction, reporting testing as an active learning experience.<sup>73</sup>

There has been limited enthusiasm for OSCE in anesthesiology programs. There are serious issues with cost, feasibility,<sup>74-77</sup> and reliability.<sup>41,45,52,62,78</sup> 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.<sup>79,80</sup> The optimum feedback format is written, if the goal is to stim-

ulate learning.<sup>81</sup> Validity of global assessment improves when structured criteria are used.<sup>23</sup> An example of structured criteria is the RIME terminology of Pangaro,<sup>2</sup> which uses descriptive terms for ascending skill levels of performance.<sup>2</sup> Without structure, the majority of strengths/weaknesses were missed by experienced IM faculty compared with assessments with a structured format.<sup>82,83</sup> The reliability issue of the "easy grader" is magnified with global ratings of observed performance.<sup>84,85</sup> Interrater reliability is low even with extensive faculty training.<sup>86</sup> Longitudinal observations using a template yield superior results compared with observations without a template.<sup>87</sup> Prestablished criteria for direct observation are valuable to identify skill levels and the need for remediation of highly technical tasks.<sup>88,89</sup> Direct observation on multiple occasions by the same observer is more reliable than observation on a single occasion.<sup>90</sup> The reliability of multiple encounter observation is better if criteria are rigid when different observers are used.<sup>91</sup>

Direct observation of anesthesiology residents for assessment of patient care was indirectly validated by Rhoton *et al.*,<sup>92</sup> 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,<sup>93</sup> especially if the training included explicit expectations.<sup>94</sup> An added advantage is the additional learning from the act of self-assessment.<sup>95</sup> Specific training for reflection improves the ultimate product in a system of self-assessment.<sup>96</sup> 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.<sup>97</sup> In a general practice setting, reflection about challenging cases combined with journaling and third-party feedback improved self-assessment skills.<sup>98</sup> Student performance on self-assessment activities matched their progress in clinical skill acquisition.<sup>99</sup> Oral surgery residents were able to accurately identify areas of skill in which they required more experience and teaching.<sup>100</sup> When initial attempts at self-assessment by residents were compared with subsequent attempts, training and

repetition resulted in improved skill.<sup>101</sup> Self-assessment may be more effective when combined with auditing and feedback for residents.<sup>102</sup> In general, trained self-assessment is harsher than faculty assessment of the same event.<sup>103</sup>

Self-assessment has not made significant inroads within anesthesiology education, although one report of self-reporting of medical errors suggested good educational merit.<sup>104</sup> A monitoring process could also have the same effect.<sup>105</sup> 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.<sup>106-108</sup> SPs can be adults or children,<sup>109-111</sup> although children as subjects have feasibility issues.<sup>109</sup> Providing the SP with a simple script makes the interaction more active in assessing consultation skills.<sup>112</sup> In a highly controlled application, SP provided an effective part of an assessment tool for patient care skills of surgical residents.<sup>113</sup> SPs also proved to be a reliable means of assessing technical skills in an EM training program.<sup>114</sup> 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.<sup>115</sup> 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.<sup>116</sup> In a medical student setting, SP performance correlated well with clinical performance.<sup>117</sup>

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,<sup>118</sup> and FM residencies because the patients can be easily recruited from continuity clinics, although faculty development effort is considerable.<sup>119,120</sup> 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.<sup>121,122</sup> The completeness of physical examination can be assessed by audit in a primary care setting.<sup>123</sup> In clinical settings, auditing is a highly effective assessment

tool with the additional advantage of encouraging the preferred clinical behavior.<sup>124</sup>

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,<sup>125,126</sup> bronchoscopy,<sup>127</sup> endoscopy,<sup>128,129</sup> laparoscopy,<sup>130</sup> lumbar puncture,<sup>131</sup> and various surgical maneuvers.<sup>132-134</sup> It has also proven effective for measurement of rapid problem-solving skills in the acute care context.<sup>135,136</sup>

Simulation has demonstrated considerable promise in anesthesiology for the teaching and assessment of the management of acute clinical crisis,<sup>137-139</sup> similar to the simulation of aviation "near-misses."<sup>140</sup> A logical extension would be the use of simulation for certification, which has been implemented for medical licensure in Italy.<sup>141</sup> In a surgical residency, simulation performance correlated well with global clinical evaluations of technical performance in the operating room.<sup>142</sup> In a medical student setting, simulation assessment was compared with global evaluation and SP performance and found to correlate well.<sup>117</sup> Assessment using simulation enhances learning in a way not achieved by didactic teaching<sup>143,144</sup> and textbooks.<sup>145</sup> Simulation may be an ideal approach to the measurement of acute care skills in anesthesiology.<sup>77</sup> Defining behavior related to critical incidents may provide a unique means of assessment of anesthesiology residents.<sup>146</sup>

### **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.<sup>147</sup> 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.<sup>148</sup> True-false questions probably should not be used.<sup>149</sup> When traditional MCQ tests were compared with OSCE short essay and extended matching for medical students, each measured a different subset of knowledge.<sup>150</sup> Using written ITEs in conjunction with some other tools may be a better approach to achieve comprehensive assessment of medical knowledge.<sup>20</sup> In a radiology residency, global performance evaluation was used to predict ITE results, with poor correlation.<sup>151</sup> ITE performance in a psychiatry residency predicted cognitive skills but did not necessarily predict clinical skills.<sup>152</sup> ITE scores did not correlate with global performance evaluations in a surgery residency.<sup>153</sup> An IM residency reviewed<sup>142</sup> multiple assessment instruments and concluded that combinations of tools were needed to achieve comprehensive assessment.<sup>45</sup> 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.<sup>153</sup>

It is not clear how well measurement of knowledge *via* MCQs translates to application of medical knowledge.<sup>154,155</sup> 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.<sup>8</sup> This was also reported in a radiology residency.<sup>154,156</sup> There was a low correlation with knowledge assessed using SPs in undergraduate medical education compared with written examination results.<sup>107</sup> Case presentation can be used for evaluation of medical knowledge with high reliability, especially if the elements presented are measured against a template.<sup>155,156</sup> 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.<sup>122</sup> For physicians in practice, self-assessment using MCQs seems to be an excellent approach to continuous professional development in a rheumatology setting.<sup>157</sup>

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.

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

#### Oral Examination

Oral examinations have a role in assessment of medical knowledge that is distinct from standardized written examinations.<sup>158</sup> Resources must be invested in faculty development to ensure reliability, because oral examinations require both questions and human examiners.<sup>159</sup> The conduct of the examination as well as the examiners must be structured to ensure standardization between candidates.<sup>160</sup> 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.<sup>161,162</sup>

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

#### 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 on-line 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 reflection.

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).<sup>167</sup> There have also been reports of assessment of evidence-based medicine programs, including audits in primary care practice,<sup>168</sup> an evidence-based medicine skills test for IM residents,<sup>169</sup> and a Web exercise.<sup>170</sup>

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.<sup>171</sup> This learning and assessment loop has been described for EM.<sup>172</sup> 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.<sup>104</sup> Self-reporting is especially relevant because physicians consistently identify the review of critical incidents and medical errors as the most significant impetus for change.<sup>173,174</sup> Self-reporting combined with peer review can be used for the assessment of PBLI for the underperforming physician.<sup>175</sup> Self-reporting with group discussion results in PBLI assessment *via* comparison feedback.<sup>176</sup>

Videotape and information management technology are excellent adjuncts to the assessment of PBLI by improving the accuracy of self-reporting.<sup>177</sup> 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.<sup>178</sup> All informatics have the potential to improve assessment of PBLI in GME.<sup>179</sup>

Self-reporting is a regular part of continuous quality improvement programs that exist in virtually every an-

esthesiology 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.<sup>78,180,181</sup> Many patient outcomes improve with excellent physician-patient interaction. These same variables can be measured by OSCE assessment<sup>182,183</sup> or video-assisted OSCE.<sup>184</sup> Effective written communication can be taught and measured in an OSCE format that focuses on written communication skills.<sup>76,185</sup> OSCE format can be used to measure language skills in international medical school graduates.<sup>186</sup>

Case presentation is a less demanding alternative to OSCE for measuring verbal communication skills.<sup>156,187</sup> Creating scripts for interacting with patients will improve communication, as long as it is observed.<sup>188</sup> Sessions where residents viewed tapes of difficult patient interviews resulted in better subsequent patient interviews compared with interviews by residents without the teaching intervention.<sup>189</sup> In an excellent example of an integrated project for teaching and assessment, Morgan and Winter<sup>190</sup> 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.<sup>191</sup> A multistep program for improved writing in medical records has been reported for a psychiatry residency.<sup>192</sup>

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

assessment of communication skills during practice oral examinations.

#### Peer Review

Properly structured peer review yields useful assessment information about communication skills for physicians in practice,<sup>193</sup> interns,<sup>194</sup> and first-year medical students.<sup>195</sup> The 360-degree review can be used to measure communication skills.<sup>196</sup>

The 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 limited 360-degree 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<sup>186,197–200</sup> and in a variety of specialties, including EM,<sup>201,202</sup> orthopedic surgery,<sup>203,204</sup> and obstetrics and gynecology<sup>205</sup> have been published. More challenging is using these general resources to create measurable endpoints that can be assessed.<sup>200</sup> Peer assessment and self-assessment,<sup>206</sup> the OSCE format,<sup>207</sup> patient feedback,<sup>208</sup> role models,<sup>209</sup> SPs,<sup>194,210,211</sup> and simulation<sup>212</sup> have been used to measure ethical behavior and professionalism.

Case-based problem resolution is also a means to measure professionalism.<sup>213</sup> Active intervention to resolve episodes of unprofessional behavior is also an effective means to assess professionalism.<sup>214</sup>

Comprehensive review in the 360-degree format has

been reported for assessment of professionalism in medical students,<sup>215</sup> for physical medicine and rehabilitation residents,<sup>216</sup> and for radiology residents.<sup>196</sup> These 360-degree evaluations yielded data, but it required considerable effort (feasibility), resulting in a limited amount of new information.<sup>217</sup> The opposite extreme has also been reported, with episodes of unprofessional behavior correlating with critical incidents.<sup>92</sup> Identifying residents with behavior issues in clinically relevant settings has been described for EM residents.<sup>218</sup>

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.<sup>219</sup> In another setting, peers were selected by the evaluatee, questionnaires were sent by mail, and the information sought was open-ended.<sup>193</sup> The unstructured format pri-



marily yielded information about communication skills, empathy, and interpersonal skills. In another report, rigid response prompts actually suppressed effective peer review, and open-ended responses were highly valuable.<sup>220</sup> All of these are assessments of the ability to practice in a healthcare system.

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

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.<sup>221</sup> Use of videotape and information handling technologies can achieve assessment of various endpoints.<sup>177</sup> 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.<sup>222-224</sup> Use of continuing medical education as a criterion for licensure is a more tangible means of defining expectations.<sup>224</sup> Combining multiple expectations may be the most effective means of ensuring practice change.<sup>225</sup> 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.<sup>226</sup>

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

### 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.<sup>228,229</sup> There are several institutions in the United Kingdom and The Netherlands leading medical education in the use of student portfolios.<sup>230-234</sup>

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.<sup>235-236</sup> 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.<sup>237-240</sup> 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.<sup>241</sup> In this process, ensuring that appropriate progress is occurring and setting learning goals that specify activities addressing areas of weakness is essential.<sup>234</sup> 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.<sup>242</sup>

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.<sup>243</sup> The technical difficulty of accumulating the data can be improved with computerization.<sup>244</sup> 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.<sup>245</sup> Some portfolio assessment projects have been reported in GME, including psychiatry<sup>246</sup> and EM.<sup>247</sup> Higher test scores as evidence of improved learning as a result of portfolio assessment have been reported in undergraduate medical education.<sup>248</sup> 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.<sup>249</sup> The use of one

portfolio process to assess all six competencies has been described in a psychiatry residency.<sup>250</sup> 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<sup>251</sup> 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<sup>251</sup> between the American Board of Anesthesiology data and the ACGME competencies, further reinforcing the optimum cycle of assessment that encourages learning.

## References

1. Kassirer JP: Pseudoaccountability. *Ann Intern Med* 2001; 134:587-90
2. Pangano L: A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med* 1999; 74:1203-7
3. O'Donnell MJ, Obenshain SS, Erdmann JB: Background essential to the proper use of results of step 1 and step 2 of the USMLE. *Acad Med* 1993; 68:734-9
4. Williams RG III: Use of NBME and USMLE examinations to evaluate medical programs. *Acad Med* 1993; 68:748-52
5. Berner ES, Brooks CM, Erdmann JB: Use of the USMLE to select residents. *Acad Med* 1993; 68:753-5

6. Slogoff S, Hughes FP: Validity of scoring “dangerous answers” on a written certification examination. *J Med Educ* 1987; 62:625-31
7. Kearney RA, Sullivan P, Skakun E: Performance on ABA-ASA in-training examination predicts success for RCPSC certification. *Can J Anesth* 2000; 47:914-8
8. Cox SM, Herbert WNP, Grosswald SJ, Carpentieri AM, Visscher HC, Laube DW: Assessment of the in-training examination in obstetrics and gynecology. *Obstet Gynecol* 1994; 84:1051-4
9. Ram P, vander Vleuten C, Rethans JJ, Schouten B, Hobma S, Grol R: Assessment in general practice: The predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. *Med Educ* 1999; 33:197-203
10. Ferland JJ, Dorval J, Levasseur L: Measuring higher cognitive levels by multiple choice questions: A myth? *Med Educ* 1987; 21:109-13
11. Hawkins RE, Sumption KF, Gaglione MM, Holmbor ES: The in-training examination in internal medicine: Resident perceptions and the lack of correlation between resident scores and faculty predictions of resident performance. *Am J Med* 1999; 106:206-10
12. Delandshere G, Petrosky AR: Capturing teachers' knowledge: Performance assessment. *Educ Reseacher* 1994; 23:11-8
13. Friedman Ben-David M: The role of assessment in expanding professional horizons. *Med Teacher* 2000; 22:9-16
14. Harden RM, Crosby JR, Davis MH, Friedman M: Outcome-based education: V. From competency to metacompetency: A model for the specification of learning outcomes. *Med Teacher* 1999; 21:546-52
15. Abraham S: Gynaecological examination: A teaching package integrating assessment with learning. *Med Educ* 1998; 32:76-81
16. Gordon MJ: Cutting the Gordian knot: A two part approach to the evaluation and professional development of resident. *Acad Med* 1997; 72:876-80
17. Ramsay PG, Wenrich MD, Carline JC, Invi TS, Larson EB, Lo Gerfo JP: Use of peer rating to evaluate physician performance. *JAMA* 1993; 269:1655-60
18. Pitts J, Coles C, Thomas P: Educational portfolios in the assessment of general practice trainees: Reliability of assessor. *Med Educ* 1999; 33:478-9
19. Dauphinee WD: Assessing clinical performance: Where do we stand and what might we expect? *JAMA* 1995; 274:741-3
20. Turnbull J, Gray J, MacFadyen J: Improving in-training evaluation programs. *J Gen Int Med* 1998; 13:317-23
21. Haber RJ, Avins AL: Do ratings on the American Board of Internal Medicine resident evaluation form detect differences in clinical competence? *J Gen Intern Med* 1994; 9:140-5
22. Borman WC: Effect of instruction to avoid halo error on reliability and validity of performance evaluation rating. *J App Psychol* 1975; 60:556-60
23. Vander Vleuten CPM, Norman GF, de Graaf E: Pitfalls in the pursuit of objectivity: Issues of reliability. *Med Educ* 1991; 25:110-8
24. Maxim BR, Dielman TE: Dimensionality, internal consistency and inter rater reliability of clinical performance ratings. *Med Educ* 1987; 27:130-7
25. Phelan S: Evaluation of the non-cognitive professional traits of medical students. *Acad Med* 1993; 68:799-803
26. Norman GR, vander Vleuten CPM, de Graaff E: Pitfalls in the pursuit of objectivity: Issues of validity, efficiency and acceptability. *Med Educ* 1991; 25:119-26
27. McGuire C: Perspectives in assessment. *Acad Med* 1993; 68:53-8
28. Hunt DD: Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. *Acad Med* 1992; 67:254-9
29. Phillips SE: Legal issues in performance assessment. *West's Educ Law Q* 1993; 2:329-58
30. Pangaro LN: Investing in descriptive evaluation: A vision for the future of assessment. *Med Teacher* 2000; 22:478-81
31. Hemmer PA, Pangaro LN: The effect of formal evaluation sessions during clinical clerkships in better identifying students with marginal fund of knowledge. *Acad Med* 1997; 72:641-3
32. 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; 75:167-73
33. Lavin B, Pangaro LN: Internship ratings as a validity measure for an evaluation system to identify inadequate clerkship performance. *Acad Med* 1998; 75:998-1002
34. Harden RM, Grant J, Buckley G, Hart IR: BEME Guide No. 1: Best evidence medical education. *Med Teacher* 1999; 21:553-62
35. Kane M: Model-based practice analysis and test specifications. *Appl Meas Educ* 1997; 10:5-18
36. Dunn MM, Wooliscroft JO: Assessment of a pattern-recognition examination in a clinical clerkship. *Acad Med* 1994; 69:683-4
37. Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE: The relationship between faculty ward evaluations, OSCE, and the ABSITE as measures of surgical intern performance. *Am J Surg* 1995; 169:414-7
38. Schwartz RW, Witzke DB, Donnelly MB, Stratton T, Blue AV, Sloan DA: Assessing resident's clinical performance: Cumulative results of a four-year study with the objective structured clinical examination. *Surgery* 1998; 124:307-12
39. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE: The objective structured clinical examination: The new gold standard for evaluating postgraduate clinical performance. *Ann Surg* 1995; 222:735-42

40. Davis MH: OSCE: The Dundee experience. *Med Teacher* 2003; 25:255-61
41. Alatif A: An examination of the examinations: The reliability of the objective structured clinical examination. *Med Teacher* 1992; 14:179-83
42. Chalabian J, Garman K, Wallace P, Dunnington G: Clinical breast evaluation skills of house officers and students. *Am Surg* 1996; 62:840-6
43. Martin JA, Regehr G, Reznick R, MacRae H, Hurnaghan J, Hutchinson C, Brown M: Objective structured assessment of technical skills (OSATS) for surgical residents. *Br J Surg* 1997; 84:273-8
44. MacRae H, Regehr G, Leadbetter W, Reznick RK: A comprehensive examination for senior surgical residents. *Am J Surg* 2000; 179:190-3
45. Hull AL, Hodder S, Berger B, Ginsberg D, Lindhein N, Quan J, Kleinhenz ME: Validity of three clinical performance assessments of internal medicine clerks. *Acad Med* 1995; 70:517-22
46. Dupras DM, Li JTC: Use of an objective structured clinical examination to determine clinical competence. *Acad Med* 1995; 70:1029-34
47. Frost GJ, Cater JI, Forsyth JC: The use of the objective structured clinical exam in pediatrics. *Med Teacher* 1986; 8:261-9
48. Martsell DG, Wolfish M, Hsu E: Reliability and validity of the objective structured clinical examination in pediatrics. *Med Educ* 1991; 25:293-9
49. Jograbchi B, Devries JM: Evaluation of clinical competence: The gap between expectation and performance. *Pediatr* 1996; 97:179-84
50. Trivino X, Vasquez A, Mena A, Lopez A, Aldunate M, Varas M, Lillo R, Wright A: Application of objective structured clinical examination (OSCE) for pediatric internship assessment in two schools of medicine. *Rev Med Chil* 2002; 130:817-24
51. Carraccio C, Englander R: The objective structured clinical examination: A step in the direction of competency-based evaluation. *Arch Pediatr Adolesc Med* 2000; 154:736-41
52. Martsell DG, Wolfish NM, Hsu E: Reliability and validity of the objective structured clinical examination in paediatrics. *Med Educ* 1991; 25:293-9
53. Hamadeh G, Lancaster C, Johnson A: Introducing the objective structured clinical examination to a family practice residency program. *Fam Med* 1993; 25:237-41
54. Sisley AC, Johnson SB, Erickson W, Fortune JB: Use of an Objective Structured Clinical Examination (OSCE) for the assessment of physician performance in the ultrasound evaluation of trauma. *J Trauma* 1999; 47:627-31
55. Sloan DA, Donnelly MB, Schwartz RW, McGrath PC, Kenady DE, Wood DP, Strodel WE: Measuring the ability of residents to manage oncologic problems. *J Surg Oncol* 1997; 64:135-42
56. Langford NJ, Landray M, Martin U, Kendall MJ, Ferner RE: Testing the practical aspects of therapeutics by objective structured clinical examination. *J Clin Pharm Ther* 2004; 29:263-6
57. Sloan DA, Donnelly MB, Schwartz RW, Felts JL, Blue AV, Strodel WE: The use of the objective structured clinical examination (OSCE) for evaluation and instruction in graduate medical education. *J Surg Res* 1996; 63:225-30
58. Karani R, Leipzig RM, Callahan EH, Thomas DC: An unfolding case with a linked objective structured clinical examination (OSCE): A curriculum in inpatient geriatric medicine. *J Am Geriatr Soc* 2004; 52:1191-8
59. Kwolek DS, Witzke DB, Bove AV, Schwartz RW, Sloan DA: Using an OSCE to assess the ability of residents to manage problems in women's health. *Acad Med* 1997; 72:548-50
60. Park RS, Chibnall JT, Blaskiewicz RJ, Furman GE, Powell JK, Mohr C: Construct validity of an objective structured clinical examination (OSCE) in psychiatry: Associations with the clinical skills examination and other indicators. *Acad Psychiatry* 2004; 28:122-8
61. McLay RN, Rodenhauser P, Anderson DS, Stanton ML, Markert RJ: Simulating a full-length psychiatric interview with a complex patient: An OSCE for medical students. *Acad Psychiatry* 2002; 26:162-7
62. Lentz GM, Mandel LS, Lee D, Gardella C, Melville J, Goff BA: Testing surgical skills of obstetric and gynecology residents in a bench laboratory setting: Validity and reliability. *Am J Obstet Gynecol* 2001; 184:1462-70
63. Rymer AT: The new MRCOG Objective Structured Clinical Examination: The examiners evaluation. *J Obstet Gynaecol* 2001; 21:103-6
64. Hassell AB, West Midlands Rheumatology Services and Training Committee: Assessment of specialist registrars in rheumatology: Experience an objective structured clinical examination (OSCE). *Rheumatology (Oxford)* 2002; 41:1323-8
65. MacRae H, Regehr G, Leadbetter W, Reznick R: A comprehensive examination for senior surgery residents. *Am J Surg* 2000; 179:190-3
66. Bann S, Datta V, Khan M, Darzi A: The surgical error examination is a novel method for objective technical knowledge assessment. *Am J Surg* 2003; 185:507-11
67. Sloan DA, Plymale MA, Donnelly MB, Schwartz RW, Edwards MJ, Blake KI: Enhancing the clinical skills of surgical residents through structured cancer education. *Ann Surg* 2004; 239:561-6
68. Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA: Ceiling effect in technical skills of surgical residents. *Am J Surg* 2004; 188:294-300
69. Zyromski NJ, Staren ED, Merrick HW: Surgery residents' perception of the Objective Structured Clinical Examination (OSCE). *Curr Surg* 2003; 60:533-7
70. Meerick HW, Nowacek GA, Boyer J, Padgett B, Francis P, Gohara SF, Staren ED: Ability of the objective structured clinical examination to differentiate surgical residents, medical students, and physician assistant students. *J Surg Res* 2002; 106:319-22
71. Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO: Assessing residents' competencies at baseline: Identifying the gaps. *Acad Med* 2004; 79:564-70
72. Wilkinson TJ, Fontaine S: Patients' global ratings of student competence: Unreliable contamination or gold standard? *Med Educ* 2002; 36:1117-21
73. Elman D, Hooks R, Tabak D, Regehr G, Freeman R: The effectiveness of unannounced standardized patients in the clinical setting as a teaching intervention. *Med Educ* 2004; 38:969-73
74. Kaufman DM, Mann KV, Muijtens AMM, vander Vleutens CPM: A comparison of standard setting procedures for an OSCE in undergraduate medical education. *Acad Med* 2000; 75:267-71
75. Newble DL, Swanson DB: Psychometric characteristics of the objective structured clinical examination. *Med Educ* 1988; 22:325-34
76. Verhoven BH, Hammerson JGHC, Scherpier AJJA: The effect on reliability of adding a separate written assessment components to an objective structured clinical examination. *Med Educ* 2000; 34:525-9
77. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A: Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *ANESTHESIOLOGY* 2003; 99:1270-80
78. Cohen R, Rothman AI, Bilan S, Ross J: Analysis of the psychometric properties of eight administrations of an objective structured clinical examination used to assess international medical graduates. *Acad Med* 1996; 71:522-4
79. Cudulka R, Emerman C, Jouriles N: Evaluation of resident performance and intensive bedside teaching during direct observation. *Acad Emerg Med* 1996; 3:345-51
80. Alnair FA: The Watched Structure Clinical Examination (WASCE) as a tool of assessment. *Saudi Med J* 2004; 25:71-4
81. Bing-You RG, Greenberg LW, Widerman BL, Smith CS: A randomized multicenter trial to improve resident teaching with written feedback. *Teach Learn Med* 1997; 9:10-13
82. Noel G, Herbers J, Caplow M, Cooper G, Pangaro L, Harvey J: How well do internal medicine faculty members evaluate the clinical skills of residents? *Ann Int Med* 1992; 117:757-65
83. Herberths J, Gordon N, Cooper G, Harvey J, Pangaro L, Weaver M: How accurate are faculty evaluations of clinical competence? *J Gen Intern Med* 1989; 4:202-8
84. Marienfeld RD, Reid JC: Six-year documentation of the easy grader in the medical clerkship setting. *J Med Educ* 1984; 59:589-91
85. Littlefield JH, DaRosa DA, Anderson KD, Bell RM, Nicholas GG, Wolfson PJ: Assessing performance in clerkships. *Acad Med* 1991; 66:516-8
86. Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor W, Brown FH, Karff M, Levey GS: The inter-rater reliability and internal consistency of a clinical evaluation exercise. *J Gen Intern Med* 1992; 7:174-9
87. Fowell SL, Southgate LT, Bligh FG: Evaluating assessment; the missing link? *Med Educ* 1999; 33:276-81
88. Holman JR, Marshall RC, Jordan B, Vogelmann L: Technical competency in flexible sigmoidoscopy. *J Am Board Fam Pract* 2001; 14:424-9
89. Paisley AM, Baldwin P, Paterson-Brown S: Feasibility, reliability and validity of a new assessment form with basic surgical trainees. *Am J Surg* 2001; 182:24-9
90. Turnbull J, MacFadyen J, van Barneveld C, Norman G: Clinical work sampling: A new approach to the problem of in-training evaluation. *J Gen Intern Med* 2000; 15:556-61
91. Brennan B, Norman GR: Use of encounter cards for evaluation of residents in obstetrics. *Acad Med* 1997; 72:543-4
92. Rhoton MF, Barnes A, Flashburg M, Ronai A, Springman S: Influence of anesthesiology residents noncognitive skills on the occurrences of critical incidents and the residents' overall performance. *Acad Med* 1991; 66:359-61
93. Abrams RG, Kelley ML: Student self-evaluation in a pediatric-operative technique course. *J Dent Educ* 1974; 38:385-91
94. Scabassi SE, Woelfel SK: Development of self-assessment skills in medical students. *Med Educ* 1984; 84:226-231
95. Arnold L, Willoughby TL, Calkins EV: Self-evaluation in undergraduate medical education: A longitudinal perspective. *J Med Educ* 1985; 60:21-8
96. Pee B, Woodman T, Fry H, Davenport E: Practice-based learning: Views on the development of a reflective learning tool. *Med Educ* 2000; 34:754-61
97. Grimes D, Bachicha J, Learman L: Teaching critical appraisal to medical students in obstetrics and gynecology. *Obstet Gynecol* 1998; 92:877-82
98. Al-Shehri A: Learning by reflection in general practice: A study report. *Educ Gen Pract* 1995; 7:237-48
99. Fitzgerald JT, White CB, Gruppen LD: A longitudinal study of self-assessment accuracy. *Med Educ* 2003; 37:645-9
100. Wanigasooriya N: Student self-assessment of essential skills in dental surgery. *Br Dent J* 2004; (suppl):11-4
101. Gordon MJ: A review of the validity and accuracy of self-assessment in health professions training. *Acad Med* 1991; 66:762-9
102. Leaf DA, Neighbor WE, Schaad D, Scott CS: A comparison of self-report and chart audit in studying resident physician assessment of cardiac risk factors. *J Gen Intern Med* 1995; 10:194-8
103. Stuart MR, Goldstein HS, Snope FC: Self-evaluation by residents in family medicine. *J Fam Pract* 1980; 10:639-42
104. Chopra V, Bovill JG, Spierdijk J, Koornneef F: Reported significant observations during anaesthesia: A prospective analysis over an 18-month period. *Br J Anaesth* 1992; 68:13-7

105. Kanfer FH: Self-monitoring: Methodological limitations and clinical applications. *J Consul Clin Psychol* 1970; 35:148-52
106. Anderson MB, Kassebaum DG: Proceedings of the AAMC's consensus conference on the issue of standardized patients in the teaching and evaluation of clinical skills. *Acad Med* 1993; 68:437-9
107. Gomez JM, Prieto L, Pujol R, Arbizu T, Vilar L, Borrell F, Roma J, Martinez-Carretero JM: Clinical skills assessment with standardized patients. *Med Educ* 1997; 31:94-8
108. Colliver JA, Swartz MH: Assessing clinical performance with standardized patients. *JAMA* 1997; 278:79-91
109. Lane JL, Ziu A, Boulet JR: A pediatric clinical skills assessment using children as standardized patients. *Arch Pediatr Adolesc Med* 1999; 153:637-44
110. Tsai TC: Using children as standardized patients for assessing clinical competence in paediatrics. *Arch Dis Child* 2004; 89:1117-20
111. Hanson M, Tiberius R, Hodges B, MacKay S, McNaughton N, Dickens S, Regehr G: Adolescent standardized patients: Method of selection and assessment of benefits and risks. *Teach Learn Med* 2002; 14:104-13
112. Mansfield F: Supervised role-play in the teaching of the process of consultation. *Med Educ* 1991; 25:485-90
113. MacRae HM, Choeh R, Regehr G, Reznick R, Burnstein M: A new assessment tool: The patient assessment and management examination. *Surgery* 1997; 122:335-44
114. Burdick WP, Friedman BM, Swisher L, Becher J, Magee D, McNamara R, Zwanger M: Reliability of performance-based clinical skill assessment of emergency medicine residents. *Acad Emerg Med* 1996; 3:1119-23
115. Stillman PL, Swanson DB, Sydney S, Stillman AE, Ebert TH, Emmel VS, Caslowitz J, Greene L, Hamolsky M, Hatem C, Levinson DJ, Levin R, Levisson G, Ley B, Morgan J, Parrino T, Robinson S, Williams J: Assessing clinical skills of residents with standardized patients. *Ann Int Med* 1986; 105:762-71
116. Ram P, vander Vleuten C, Rethans J, Grol R, Aretz K: Assessment of practicing family physicians: Comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. *Acad Med* 1999; 74:62-9
117. Edelstein R, Reid RD, Usatine R, Wilkes M: A Comparative study of measures to evaluate medical students' performance. *Acad Med* 2000; 75:825-33
118. Hall MJ, Adamo G, McCurry L, Lacy T, Waits W, Chow J, Rawn L, Ursano RJ: Use of standardized patients to enhance a psychiatry clerkship. *Acad Med* 2004; 79:28-31
119. Feeley TH, Manyon AT, Servoss TJ, Panzarella KJ: Toward validation of an assessment tool designed to measure medical students' integration of scientific knowledge and clinical communication skills. *Eval Health Prof* 2003; 26:222-33
120. Amano H, Sano T, Gotoh K, Kakuta S, Suganuma T, Kimura Y, Tsukasaki H, Miyashita H, Okano T, Goto N, Saeki H: Strategies for training standardized patient instructors for a competency exam. *J Dent Educ* 2004; 68:1104-11
121. Holmboe E, Scranton R, Sumption K, Hawkins R: Effect of medical record audit and feedback on residents' compliance with preventative health care guidelines. *Acad Med* 1998; 73:901-3
122. Ramsdell JW, Berry CC: Evaluation of general and traditional internal medicine residencies utilizing a medical records audit based on educational objectives. *Med Care* 1983; 21:1144-53
123. Ognibene AJ, Jarjoura DG, Illera VA, Blend DA, Cugino AE, Whittier F: Using chart reviews to assess residents' performances of components of physical examinations: A pilot study. *Acad Med* 1994; 69:583-7
124. Wainwright JR, Sullivan FM, Morrison JM, MacNaughton RJ, McConnadrie A: Audit encourages an evidence-based approach to medical practice. *Med Educ* 1999; 33:907-14
125. Chapman DM, Rhee JK, Marx JA, Honigman B, Panacek EA, Martinez D, Brofeldt BT, Cavanaugh SH: Open thoracotomy procedural competency: Validity study of teaching and assessment modalities. *Ann Emerg Med* 1996; 28:641-7
126. Chapman DM, Marx JA, Honigman B, Rosen P, Cavanaugh SH: Emergency thoracotomy: Comparison of medical student, resident, and faculty performances on written, computer, and animal-model assessments. *Acad Emerg Med* 1994; 1:373-81
127. Crawford SW, Colt HG: Virtual reality and written assessments are of potential value to determine knowledge and skill in flexible bronchoscopy. *Respiration* 2004; 71:269-75
128. Macmillan AIM, Cushieri A: Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: Predictive and concurrent validity. *Am J Surg* 1999; 177:274-7
129. Sedlack RE, Kolars JC, Alexander JA: Computer simulation training enhances patient comfort during endoscopy. *Clin Gastroenterol Hepatol* 2004; 2:348-52
130. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A: Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: Results from randomized controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform* 1998; 50:124-30
131. Moorthy K, Jiwani M, Shah J, Bello F, Munz Y, Darzi A: Validation of a web-based training tool for lumbar puncture. *Stud Health Technol Inform* 2003; 94:219-25
132. Paisley AM, Baldwin PJ, Paterson-Brown S: Validity of surgical simulation for the assessment of operative skill. *Br J Surg* 2001; 88:1525-32
133. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W: Testing technical skill *via* an innovative "Bench Station" examination. *Am J Surg* 1996; 172:226-30
134. Szalay D, MacRae H, Regehr G, Reznick R: Using operative outcome to assess technical skill. *Am J Surg* 2000; 180:234-7
135. Stringer J, Moreno EA: Computers in simulation: Applications in medical education. A computer literacy requirement for medical students. *Acad Med* 1996; 71:522-4
136. McLaughlin SA, Doezema D, Sklar DP: Human simulation in emergency medicine training: A model curriculum. *Acad Emerg Med* 2002; 9:1310-8
137. Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R: Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *ANESTHESIOLOGY* 1998; 89:8-18
138. Murray DJ, Boulet JR, Kras JF, Woodhouse JA, Cox T, McAllister JD: Acute care skills in anesthesia practice: A simulation-based resident performance assessment. *ANESTHESIOLOGY* 2004; 101:1084-95
139. Schwid HA, Rooke GA, Carline J, Steadman RH, Murray WB, Olympio M, Tarver S, Steckner K, Wetstone S, Anesthesia Simulator Research Consortium: Evaluation of anesthesia residents using mannequin-based simulation: A multi-institutional study. *ANESTHESIOLOGY* 2002; 97:1434-44
140. Helmreich RL, Davies JM: Anaesthetic simulation and lessons to be learned from aviation. *Can J Anaesth* 1997; 44:907-12
141. Guagnano MT, Merlitti D, Manigrasso MR, Pace-Palitti V, Sensi S: New medical licensing examination using computer-based case simulations and standardized patients. *Acad Med* 2002; 77:87-90
142. Scott D, Valentine J, Bergen P, Rege R, Laycock R, Tesfay S, Jones D: Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. *Surgery* 2000; 128:613-22
143. McMahon GT, Monaghan C, Falchuk K, Gordon JA, Alexander EK: A simulator-based curriculum to promote comparative and reflective analysis in an internal medicine clerkship. *Acad Med* 2005; 80:84-9
144. Hudson JN: Computer-aided learning in the real world of medical education: Does the quality of interaction with the computer affect student learning? *Med Educ* 2004; 28:887
145. Qayumi AK, Kurihara Y, Imai M, Pachev G, Seo H, Hoshino Y, Cheifet R, Matsuura K, Momo M, Saleem M, Lara-Guerra H, Miki Y, Kariya Y: Comparison of computer-assisted instruction (CAI) *versus* traditional textbook methods for training in abdominal examination (Japanese experience). *Med Educ* 2004; 38:1080-8
146. Altmair EM, From RP, Pearson KS, Gorbatenko-Roth KG, Ugolini KA: A prospective study to select and evaluate anesthesiology resident: Phase 1, the critical incident technique. *J Clin Anesth* 1997; 9:629-36
147. Norcini JJ, Swanson DB, Grosso LF, Shea JA, Webster GD: A comparison of knowledge, synthesis and clinical judgment: Multiple choice questions in the assessment of physician competence. *Eval Health Professional* 1984; 7:485-500
148. vander Vleuten CPM: Validity of final examinations in undergraduate medical training. *BMJ* 2000; 321:1217-9
149. Anderson J: Multiple choice questions revisited. *Med Teacher* 2004; 26:110-3
150. Wass V, McGibbon D, vander Vleuten G: Composite undergraduate clinical examinations: How should the components be combined to maximize reliability. *Med Educ* 2001; 35:326-30
151. Wise S, Stagg L, Szucs R, Gay S, Mauger D, Hartman D: Assessment of resident knowledge: Subjective assessment *versus* performance on the ACR In-Training Examination. *Radiol Educ* 1999; 6:66-71
152. Webb L, Sexson S, Scully J, Reynolds C, Shore M: Training directors' opinions about the psychiatry resident In-Training Examination (PRITE). *Am J Psychiatry* 1992; 149:521-4
153. Schwartz R, Donnelly M, Sloan D, Johnson S, Strodel W: Assessing senior residents' knowledge and performance: An integrated evaluation program. *Surgery* 1994; 116:634-40
154. Adusumilli S, Cohan RH, Korobkin M, Fitzgerald JT, Oh MS: Correlation between radiology resident rotation performance and examination scores. *Acad Radiol* 2000; 7:920-6
155. Schuwirth LWT, Southgate L, Page GG, Paget NS, Lescop JMJ, Lew SR, Wade WB, Baron-Maldonado M: When enough is enough: A conceptual basis for fair and defensible practice performance assessment. *Med Educ* 2002; 36:925-30
156. Blanc CE, Calhoun JG: Objectively evaluating student case presentations. *Invest Radiol* 1985; 20:121-3
157. Beyeler C, Westkamper R, Villiger PM, Aeschlimann A: Self assessment in continuous professional development: A valuable tool for individual physicians and scientific societies. *Ann Rheum Dis* 2004; 63:1684-6
158. Eagle CJ, Martineau R, Hamilton K: The oral examination in anesthetic resident evaluation. *Can J Anaesth* 1993; 40:947-53
159. Des Marchais JE, Jean P: Effects of examiner training on open-ended, higher taxonomic level questioning in oral examinations. *Teach Learn Med* 1993; 5:24-8
160. Ferron D: Guidelines for conduct of oral examinations. *Ann R Coll Physicians Surg Canada* 1998; 31:28-31
161. James FM: Oral practice examinations: Are they worth it? *ANESTHESIOLOGY* 1999; 91:4-6
162. Pope WDB: Anaesthesia oral examination. *Can J Anaesth* 1993; 40:907-10

163. Quattlebaum TG, Darden PM, Sperry JB: In-training examinations as predictors of resident clinical performance. *Pediatrics* 1989; 84:165-72
164. Kearney Ra, Puchalski SA, Yang HYH, Skakun EN: The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Can J Anesth* 2002;49:232-6
165. Schubert A, Tetzlaff JE, Tan M, Ryckman JV, Mascha E: Consistency, inter-rater reliability and validity of 441 consecutive mock oral examinations in anesthesiology. *ANESTHESIOLOGY* 1999; 91:288-98
166. Wade TP, Andrus CH, Kaminski DL: Evaluations of surgery resident performance correlate with success in board examinations. *Surgery* 1993; 113:644-8
167. Baldwin DC, Daugherty SR, Rowley BD: Unethical and unprofessional conduct observed by residents during their first year of training. *Acad Med* 1998; 73:1195-200
168. Sweeney K: How can evidence-based medicine help patients in general practice? *Fam Pract* 1996; 13:489-90
169. Smith CA, Ganschow PS, Reilly BM, Evans AT, McNutt RA, Osei A, Saquib M, Surabhi S, Yadar S: Teaching residents evidence-based medicine skills. *J Gen Intern Med* 2000; 15:710-5
170. Lloyd FJ, Reyna VF: A web exercise in evidence-based medicine using cognitive theory. *J Gen Intern Med* 2001; 16:94-9
171. Connor MP, Bynoe AG, Redfern N, Pokora J, Clarke J: Developing senior doctors as mentors: A form of continuing professional development. Report of an initiative to develop a network of senior doctors as mentors: 1994-99. *Med Educ* 2000; 34:747-53
172. Galicia AR, Klima RR, Date ES: Mentorship in physical medicine and rehabilitation residencies. *Am J Phys Med Rehabil* 1997; 76:268-75
173. Allery LA, Owen PA, Robling MR: Why general practitioners and consultants change their clinical practice: A critical incident study. *BMJ* 1997; 314:870-4
174. Campbell C, Parboosingh J, Gondocz T, Babitskaya G, Pham B: Learning, change and practicing physicians: A study of the factors that influence physicians' commitments to change their practices using learning diaries. *Acad Med* 1999; 74:S34-36
175. Southgate L, Cox J, David T, Hatch D, Howes A, Johnson N, Jolly B, Macdonald E, McAvoy P, McCrorie P, Turner J: The assessment of poorly performing doctors: The development of the assessment programmes for the General Medical Councils' performance procedures. *Med Educ* 2001; 35:2-8
176. Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Barnett GO: Improving physician performance through peer comparison feedback. *Med Care* 1984; 22:527-34
177. Barnes BE: Creating the practice-learning environment: Using information technology to support a new model of continuing medical education. *Acad Med* 1998; 73:278-81
178. Garvin R, Otto F, McRae D: Using handheld computers to document family practice resident procedure experience. *Fam Med* 2000; 32:115-8
179. Jerant AF: Training residents in medical informatics. *Fam Med* 1993; 31:465-72
180. Hodges B, Trunbull J, Cohen R, Bienenstock A, Norman G: Evaluation of communication skills in an objective structured clinical examination format: Reliability and generalizability. *Med Educ* 1996; 30:38-44
181. Yudkowsky R, Alseidi A, Cintron J: Beyond fulfilling the core competencies: An objective structured clinical examination to assess communication and interpersonal skills in a surgical residency. *Curr Surg* 2004; 61:499-503
182. Stewart M, Brown JB, Boon H, Galajda J, Meredith L, Sangster M: Evidence on patient-doctor communication. *Cancer Prev Control* 1999; 3:25-30
183. Stewart MA: Effective physician-patient communication and health outcomes: A review. *Can Med Assoc J* 1995; 152:1423-32
184. Humphris GM, Kineu S: The objective structured video exam for assessment of communication skills. *Med Educ* 2000; 34:939-45
185. Keely E, Myers K, Dojeiji S: Can written communication skills be tested in an objective structured clinical examination format? *Acad Med* 2002; 77:82-6
186. Rothman DJ: Medical professionalism: Focusing on the real issues. *N Engl J Med* 2000; 342:1284-6
187. Roth CS, Watson KV, Harris IB: A communication assessment and skill-building exercise (CASE) for first-year residents. *Acad Med* 2002; 77:746-7
188. Boulton M, Griffiths J, Hall D, McIntyre M, Oliver B, Woodward J: Improving communication: A practical programme for teaching trainees about communication issues in the general practice consultation. *Med Educ* 1984; 18:269-74
189. Roter DL, Hall JA, Kern DE, Barber LR, Cole KA, Roca RP: Improving physicians' interview skills and reducing patients' emotional distress. *Arch Intern Med* 1995; 155:1877-84
190. Morgan ER, Winter RJ: Teaching communication skills: An essential part of residency training. *Arch Pediatr Adolesc Med* 1996; 150:638-42
191. Smith RC, Lyles JS, Mettler J, Stoeffelmayer BE, Van Egeren LF, Marshall AA, Gardner JC: The effectiveness of intensive training for residents in interviewing. *Ann Int Med* 1998; 128:118-26
192. Tinsley JA: An educational intervention to improve residents' inpatient charting. *Acad Psychiatry* 2004; 28:136-9
193. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP: Use of peer ratings to evaluate physician performance. *JAMA* 1993; 269:1655-60
194. Kegel-Flom P: Predicting supervisor, peer, and self ratings of intern performance. *J Med Educ* 1975; 50:812-5
195. Rudy DW, Fejfar MC, Griffith CH III, Wilson JF: Self-and peer assessment in a first-year communication and interviewing course. *Eval Health Prof* 2001; 24:436-45
196. Wood J, Collins J, Burnside ES, Albanese MA, Propeck PA, Kelcz F, Spille JM, Schmaltz LM: Patient, faculty, and self-assessment of radiology resident performance: A 360-degree method of measuring professionalism and interpersonal/communication skills. *Acad Radiol* 2004; 11:931-9
197. ABIM Foundation, ACP-ASIM Foundation, European Federation of Internal Medicine: Medical professionalism in the new millennium: A physician charter. *Ann Intern Med* 2002; 136:243-6
198. Pellegrino ED: Medical professionalism: Can it, should it survive? *J Am Board Fam Pract* 2000; 12:147-9
199. Pellegrino ED, Relman AS: Professional medical associations: Ethical and practical guidelines. *JAMA* 1999; 282:984-6
200. Swick HM: Toward a normative definition of medical professionalism. *Acad Med* 2000; 75:612-6
201. Adams J, Schmidt T, Sanders A, Larkin GL, Knopp R: Professionalism in emergency medicine. *Acad Emerg Med* 1998; 5:1193-9
202. Larkin GL, Binder L, Houry D, Adams J: Defining and evaluating professionalism: A core competency for graduate emergency medicine education. *Acad Emerg Med* 2002; 9:1249-56
203. Baldwin JrDC, Bunch WH: Moral reasoning, professionalism, and the teaching of ethics to Orthopaedic Surgeons. *Clin Ortho Rel Research* 2000; 378:97-103
204. Rowley BD, Baldwin DC, Bay RC, Cannula M: Can professional values be taught? A look at residency training. *Clin Ortho Rel Research* 2000; 378:110-4
205. Fries MH: Professionalism in obstetrics-gynecology residency education: The view of program directors. *Obstet Gynecol* 2000; 95:314-6
206. Asch E, Saltzberg D, Kaiser S: Reinforcement of self-directed learning and the development of professional attitudes through peer-and self-assessment. *Acad Med* 1998; 73:575
207. Singer PA, Robb A, Cohen R, Norman G, Turnbull J: Performance-based assessment of clinical ethics using an objective structured clinical examination. *Acad Med* 1996; 74:495-8
208. Laine C, Davidoff F: Patient-centered medicine: A professional evolution. *JAMA* 1996; 275:152-6
209. Wright SM, Kern DE, Kolodner K, Howard DM, Brancati LF: Attributes of excellent attending-physician role models. *N Engl J Med* 1998; 339:1986-93
210. Smith SR, Balint JA, Krause KC, Moore-West M, Viles PH: Performance-based assessment of moral reasoning and ethical judgment among medical students. *Acad Med* 1994; 69:381-6
211. Stern DT, Frohna AZ, Gruppen LD: The prediction of professional behaviour. *Med Educ* 2005; 39:75-82
212. Gisondi MA, Smith-Coggins R, Harter PM, Soltysik RC, Yarnold PR: Assessment of resident professionalism using high-fidelity simulation of ethical dilemmas. *Acad Emerg Med* 2004; 11:931-7
213. Ginsburg S, Regehr G, Hatala R: Context, conflict and resolution: A new conceptual framework for evaluating professionalism. *Acad Med* 2000; 75:86-10
214. Papadakis MA, Osborn EHS, Cooke M, Healy K: A strategy for the detection and evaluation of unprofessional behavior in medical students. *Acad Med* 1999; 74:980-90
215. Rees C, Shepherd M: The acceptability of 360-degree judgments as a method of assessing undergraduate medical students' personal and professional behaviours. *Med Educ* 2005; 39:49-57
216. Musick DW, McDowell SM, Clark N, Salcido R: Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. *Am J Phys Med Rehabil* 2003; 82:394-402
217. Weigelt JA, Brasel KJ, Bragg D, Simpson D: The 360-degree evaluation: Increased work with little return? *Curr Surg* 2004; 61:616-26
218. Marco CA: Ethics seminars: Teaching professionalism to "problem" residents. *Acad Emerg Med* 2002; 9:1001-6
219. Albanese M, Prucha C, Barnett JH: Student attitudes in evaluating courses, faculty, and peers: Labeling each response option and the direction of the positive options impacts student course ratings. *Acad Med* 1997; 27:54-6
220. Thomas PA, Gebo KA, Hellmann DB: A pilot study of peer review in residency training. *J Gen Intern Med* 1999; 14:551-4
221. Berwick DM: Continuous improvement as an ideal in health care. *N Engl J Med* 1989; 320:53-56
222. Frankford DM: Community, professionals, and participation. *J Health Polit Policy Law* 1997; 22:101-4
223. Cruess SR, Cruess RL: Professionalism: A contract between medicine and society. *CMAJ* 2000; 162:668-71
224. Fox RD, Bennett NL: Learning and change; implications for continuing medical education. *BMJ* 1998; 316:466-8
225. Greco PJ, Eisenberg JM: Changing physicians' practices. *N Engl J Med* 1993; 329:1271-4
226. Davis D, O'Brien M, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A: Impact of formal continuing medical education. *JAMA* 1999; 282:867-74
227. Rethans J, Van Leeuwe Y, Drop R, vander Vlieten C, Sturmans F: Competence and performance: Two different concepts in the assessment of quality of medical care. *Fam Pract* 1990; 7:168-74
228. Jasper MA: The potential of the professional portfolio for nursing. *J Clin Nurs* 1995; 4:249-55

229. Koretz D: Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assess Educ Principles Policy Pract* 1998; 5:309-35
230. Driessen EW, Van Tartwizk J, Vermunt JD, vander Vleuten CPM: Use of portfolios in early undergraduate medical training. *Med Teacher* 2003; 25:18-23
231. Challis M: AMEE Medical Education Guide No. 11 (revised): Portfolio-based learning and assessment in medical education. *Med Teacher* 1999; 21:370-86
232. Challis M: Portfolios and assessment: Meeting the challenge. *Med Teacher* 2001; 23:27-31
233. Friedman Ben-David M, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ: AEE Medical Education Guide No. 24: Portfolios as a method of student assessment. *Med Teacher* 2001; 23:535-51
234. Snadden D, Thomas M: The use of portfolio learning in medical education. *Med Teacher* 1998; 20:192-9
235. Friedman Ben-David M: The role of assessment in expanding professional horizons. *Med Teacher* 2000; 22:27-32
236. Friedman Ben-David M: AMEE Guide No. 14: Outcome-based education: Part 3. Assessment in outcome-based education. *Med Teacher* 1999; 21:33-6
237. Schon DA: *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning Professions*. San Francisco, Jossey-Bass, 1987
238. Murdock-Eaton D: Reflective practice skills in undergraduates. *Acad Med* 2002; 77:734
239. Jensen GM, Saylor C: Portfolios and professional development in the health professions. *Eval Health Professions* 1994; 17:344-57
240. Jarvinen A, Kohonen V: Promoting professional development in higher education through portfolio assessment. *Assess Eval Higher Educ* 1995; 20:25-32
241. Gordon J: Assessing student's personal and professional development using portfolios and interviews. *Med Educ* 2003; 37:335-40
242. Davis MH, Friedman Ben-David M, Harden RM, Howie P, Ker C, Mcghee C, Pippard MJ, Snadden D: Portfolio assessment. *Med Teacher* 2001; 23:357-66
243. Scholes J, Webb C, Gray M, Endacott R, Miller C, Jasper M, McMullan M: Making portfolios work in practice. *J Adv Nurs* 2004; 46:595-603
244. Parboosingh J: Learning portfolios: Potential to assist health professionals with self-directed learning. *J Cont Ed Health Prof* 1996; 16:75-81
245. Herman JL, Winters L: Portfolio research: A slim collection. *Educ Leadership* 1994; 10:48-55
246. O'Sullivan PS, Cogbill KK, McClain T, Reckase MD, Clardy JA: Portfolios as a novel approach for residency evaluation. *Acad Psych* 2002; 26:173-9
247. O'Sullivan P, Greene C: Portfolios: Possibilities for addressing emergency medicine resident competencies. *Acad Emerg Med* 2002; 9:1305-9
248. Finlay IG, Maughan TS, Webster DJT: A randomized controlled study of portfolio learning in undergraduate cancer education. *Med Educ* 1998; 32:172-6
249. O'Sullivan PS, Reackase MD, McClain T, Savidge MA, Clardy JA: Demonstration of portfolios to assess competency of residents. *Adv Health Sci Educ Theory Pract* 2004; 9:309-23
250. Jarvis RM, O'Sullivan PS, McClain T, Clardy JA: Can one portfolio measure the six ACGME general competencies? *Acad Psych* 2004; 28:190-6
251. Rose SH, Burkle CM: Accreditation Council for Graduate Medical Education competencies and the American Board of Anesthesiology clinical competence committee: A comparison. *Anesth Analg* 2006; 102:212-6