Assessment in Graduate Medical Education:

A Primer for Pediatric Program Directors

Assessment in Graduate Medical Education: A Primer for Pediatric Program Directors

Program Directors Committee of the American Board of Pediatrics

Project Leaders:
Carol L. Carraccio, MD, MA
Patricia J. Hicks, MD

Contributors:
Ann E. Burke, MD
M. Douglas Jones, Jr., MD
Stephen Ludwig, MD
Gail A. McGuinness, MD
Julia A. McMillan, MD
Richard P. Shugerman, MD
Suzanne K. Woods, MD

Alan Schwartz, PhD, Editor

Suggested citation:

Schwartz A, editor. 2011. Assessment in Graduate Medical Education: A Primer for Pediatric Program Directors. Chapel Hill, NC: American Board of Pediatrics.

Assessment in Graduate Medical Education: A Primer for Pediatric Program Directors

Contents

Part I: Introduction to Assessment Principles and Techniques

Carol L. Carraccio, MD, MA, Patricia J. Hicks, MD, and Alan Schwartz, PhD

At the cusp of the millennium, The Accreditation Council for Graduate Medical Education (ACGME) presented professionals in Graduate Medical Education (GME) with the challenge of implementing competency-based medical education with a focus on six broad and partially-overlapping domains. Although developing various instructional approaches to meaningfully incorporate these competencies into the GME curriculum has not been easy, assessment of the competencies has presented the most daunting task. To develop methods of assessment that provide a sound basis for making judgments about the progression of residents, program directors need to partner with medical educators with expertise in assessment methodology. This partnership is critically important because it pairs experts in the methodology of assessment with program directors who teach and assess learners in the context of delivering care. In other words, it grounds meaningful assessment in real world practice.

The challenge of improving assessment in medical education brings many great opportunities for program directors. Most importantly, we have the opportunity to advance the field of assessment in medical education itself. Despite the difficulties and barriers this challenge presents, the rewards of our work as program directors rest on our contribution to the professional formation of physicians-in-training through informed assessment and feedback with the intent of helping them to continually improve. With meaningful individual performance assessment, program directors can also aggregate outcomes as metrics to guide program evaluation and improvement. The impact of assessment is powerful in driving change and pushing for transformation.

If we want to have impact—to bring about transformation in medical education—we need to be learners as well as teachers. Medical school and residency did not provide us with the skills needed to be medical educators. We need to embrace our own continued learning so that we can live up to the work with which we have been entrusted. As academic faculty, program directors should seek evidence, through medical education research, that their work achieves the desired outcomes. Scholarship within medical education spans a wide array of domains, such as curriculum development, instructional methods, and constructive and effective feedback—all areas where assessment of outcomes is critical. Development of methods and tools to assess resident competence is a ripe area for such scholarship and requires us to rigorously study the evidence of the validity and utility of the results they produce. Medical education research provides opportunities for regional and national scholarly presentations as well as peer-reviewed publications. In short, improving the quality of assessments in our residency programs is not only beneficial to our residents and their patients, but also provides opportunities for both personal satisfaction and advancing the field.

The Primer is intended to be a stimulus for future workshops at our various program director, clerkship director, and other professional society meetings. But our responsibility does not end here. Once we ourselves have learned, we must transmit our knowledge through faculty development at our own institutions, without which we will never have the meaningful assessment of learners that is our vision and this Primer's goal.

We offer this Primer as a way to begin conversations around the complex problem of assessment in medical education and training. In writing this Primer, program directors shared their many years of experience with different levels of learners, and then invited an assessment expert, Dr. Alan Schwartz, to join

and guide the endeavor. Each chapter focuses on the principles of assessment within the context of familiar assessment challenges that program and clerkship directors and other faculty involved in medical education face on a routine basis. At the end of each chapter, an annotated bibliography of the most relevant articles is presented to assist you with further reading. We also include a more extensive bibliography for more seasoned educators who wish to go into greater depth. The authors purposefully tried to limit the use of educational jargon; however, familiarity with the terms that we include in *italics* within the text of the chapters as well as the glossary will enable you to approach the medical education literature with a greater understanding. We hope that the result is a user-friendly document that invites you to further explore in more depth the principles and practices of assessment where we have only exposed the surface.

The primer is divided into two parts. Part I focuses on the foundation of assessment principles and methods of assessment, faculty development and program evaluation. In Part II, we apply theory to practice by illustrating these principles and methods in the context of each of the ACGME domains of competence.

We are grateful for the partnership role that the membership of the Association of Pediatric Program Directors (APPD) has played in the development of a direction for this primer. Their needs established the content focus for the primer and their grounding in the trenches helped us to develop what we hope is a meaningful and practical guide to competency-based assessment of learners.

Our vision for the future is that this primer will be step one in an ongoing process of program director development. The partnership with APPD provides a mechanism for feedback as to needed next steps for continuing our professional development.

1. Measurement Principles in Medical Education

Validity, Reliability, and Utility, Oh My!

Ann Burke, MD

"Assessment in Medical Education addresses complex competencies and thus requires quantitative and qualitative information from different sources as well as professional judgment."[1]

Rationale

As program directors, we strive to assess our trainees in a thoughtful and fair manner, both for their benefit and that of the children who rely on their competence. To be able to select appropriate assessment tools that measure aspects of performance that are meaningful, one must have an informed understanding of strengths and weaknesses of various assessment tools and techniques. Further, it is helpful to understand concepts about measurement that have practical applications in the training of future pediatricians. Key concepts that need to be considered when assessing competence are *validity*, *reliability*, and the *utility* or usefulness of a given tool when it is used for purposes of assessment.[1,2] There are multitudes of other statistical concepts and definitions that are important in activities such as analyzing research data and developing and validating assessment tools,[3,4] however this chapter will not cover those topics. Practical aspects and considerations about resident assessment should be understood to enable the program director to use and interpret assessment data in meaningful and cogent ways.[5] This is particularly important because outcomes for the learner and our patients depend on our ability to meaningfully and accurately interpret the data that we glean from our assessment tools. So, although there is much debate in the medical education literature regarding how best to assess competencies, as mentioned in Part II of this primer, several strategies can be used to involve multiple raters from multiple venues (*triangulation*), to combine assessment data, and utilize *assessment methods* in a manner that accurately tells us "how our resident is doing".

Goals

1. Become familiar and comfortable with the necessary, basic concepts of measurement as they relate to resident performance.
2. Be able to apply the concepts of validity, reliability and utility to determine the usefulness of various assessment tools and strategies.
3. Understand the strengths and limitations of assessment tools and how they can be used to provide a practical approach to assessing resident competence.

Case Example

It is Spring and that time of year when the assessment and promotions committee meets at your institution to discuss resident progression to the next level. In reviewing the files before the meeting, you become particularly worried by an intern who has some discrepant reports. Her in-training examination (ITE) score is average, she scored above her peers for the Objective Structured Clinical Examination (OSCE), her rotation "global evaluation" from a NICU attending was rather negative, and the rest of her evalua-

tions from faculty are average. She has a number of complimentary nursing evaluations. However, there is an assessment from a floor nurse that is below average. How are you to know how to weigh all of this information and make any kind of reasonable, informed decision about your intern? At the meeting, most of the faculty who have worked with her report that she is quiet, performs slightly above average, and should be promoted. One faculty member, however, says he is concerned about the intern's ability to synthesize information to formulate a reasonable assessment, and is quite adamant that "she cannot supervise others until she demonstrates this skill."

Points for Consideration

How do I know which of these assessments of my resident are valid? What does validity mean?

A *assessment instrument* is valid if it actually measures what it is supposed to measure. Evidence for validity lies on a continuum; that is, assessments are not valid or invalid, rather there is spectrum of evidence that can be used to support validity that ranges from weak to strong. In summary:

- Validity refers to evidence that supports or refutes the meaning or interpretation of assessment results.
- Assessments are not valid or invalid; rather the scores or outcomes of assessment have more or less evidence to support a specific interpretation. [5]
- Multiple sources of strong evidence are more supportive than fewer or weaker sources.

There are five sources of evidence that can be used to support validity:

Content: Does the assessment cover the areas it intends to? Factors that influence this are: (1) aligning the test or the observation to the learning objectives for the course or the rotation (*blueprinting*) and (2) adequate and representative sampling of the knowledge and/or skills being observed and judged. The resident in the case may be receiving discrepant assessments because one or more of the assessments may not be assessing what is supposed to be assessed. For example, the NICU attending may not like residents to be quiet and shy, and may be assessing demeanor rather than patient care abilities.

Response Process: This term refers to the evidence of data integrity and quality control of score reporting and accuracy. Is the instrument's result accurately reported? Is the form of assessment familiar to the learner? For example, is there a system that assures the correct resident name is on the assessment tool, or are faculty mixing residents up? Did the resident know she was being assessed by nurses? Do residents understand how to approach an OSCE encounter?

Internal Structure: This refers to reliability, and issues with rater variability fall into this category. In the case, is it possible that the raters had no guidance on how to use the assessment instrument? If the raters are not standardized, then the results of the observations may not be measuring what should be assessed. Powerful approaches to standardization include the development of a uniform rating tool, and the training of raters through applying the tool to a common observation, such as a videotaped encounter. Assessing how the raters score the items and their agreement may give insight into rating abilities and tendencies of faculty.

Relationship to other Variables: How well does the assessment tool correlate with the outcome of another measure of the same skill? For example, does the resident who scores highly on an OSCE station that is purported to assess communication skills also receive above average assessments in clinically observed interactions? Of note, positive correlation with other tools thought to assess the same skill are hard to come by in medicine because we lack gold standards for most skills.

Consequences: This considers the impact of scores on learners—whether or not the assessment is high-stakes or low-stakes. As an example, did the resident think that if she did poorly in the NICU rotation she would fail the whole year, and thus suffer from performance anxiety during observations? Conversely, did she think that the NICU rotation was unimportant?

There are many explanations of why there are discrepant "scores" on various instruments completed by various raters!

How do I know which assessment tools are worthwhile, and which ones are not? What poses a threat to validity?

To make useful assessments of residents we should be aware of two "threats to validity".[7] The first is inadequate sampling in the area or domain that you want to assess (referred to in the educational literature as *construct underrepresentation*). An example would be that the resident did "well" on the OSCE exercise, but there was not enough variety of cases to fully sample her knowledge, skills, and attitudes. The second threat to validity is error that systematically, rather than randomly, interferes with your ability to meaningfully interpret scores or ratings (referred to in the educational literature as *construct-irrelevant variance*). For example, consider a disgruntled nurse who consistently gives poor ratings on resident assessments. If you weighed the faulty, less valid assessment from that nurse more heavily than all of the other more valid assessments, you would be threatening the validity of your assessment system. Downing and Haladyna[7] provide examples of these threats in three areas of medical assessment: written test questions, such as *multiple choice questions* and *in-training exams*; performance examinations, such as *OSCEs*, *simulated patients,* and high fidelity training in simulation centers; and ratings of clinical performance such as observations of histories and physical exams by faculty members.

In written test questions, construct underrepresentation occurs when the test items do not adequately sample the domain of knowledge, either because there are not enough test items or because the items are not a good representation of the knowledge to be assessed. Construct-irrelevant variance results when items are too difficult or easy to provide meaningful scores, when learners can cheat or guess the test answers, or when passing scores are set in a way that does not appropriately discriminate between successful and unsuccessful learning.

In performance examinations, construct underrepresentation occurs when cases or stations are too few or do not adequately represent the skill to be assessed. Construct-irrelevant variance is introduced by unstandardized raters, poorly designed checklists, unreliable measurement, cases that are too difficult or easy, and passing scores set inappropriately.

In clinical ratings, construct underrepresentation occurs when too few observers or too few ratings are used to assess a resident's performance, or performance is not observed under a representative range of patients. Construct-irrelevant variance is introduced by poorly designed rating forms, poorly trained observers, and systematic bias in observers' use of the rating form (such as leniency or halo effects).

How do I address the variability of faculty when they assess my residents' knowledge, skills, and attitudes?

Variability in ratings—even ratings of the same videotaped encounter—can result from differences in what the raters are rating, differences in how the raters are using the rating instrument, and random "noise". For example, a rater may ignore the trait domains that are supposed to be rated and treat all of the traits as one. This behavior is seen when a resident does poorly on a number of physical exam skills, and yet receives a "good rating" because the attending likes her personality; this is a *halo effect*.[10] On the other hand, raters may all be focusing on the same trait, but using the rating instrument differently. The terms

leniency and *severity* describe systematic ratings that are too "easy" or too "hard" on trainees. Raters may also exhibit *central tendency* by limiting their ratings to values near the midpoint of the rating scale and eschewing extreme ratings. The effect of leniency, severity, and central tendency is to reduce the ability of the rater to discriminate among levels of performance; the reduced range of rating also limits the degree to which the ratings can be demonstrated to be consistent with those of other raters.

A number of ways to minimize these threats to the validity of assessments include:

- Increase the number of clinical observations performed. More ratings of clinical performance, even unsystematic and unstandardized, may be as effective as and more practical than simulated patient approaches. Between 7 and 11 independent ratings are probably needed to produce evidence of validity using this method.[9]
- Develop a *faculty development* program to train raters to diminish severity, leniency, central tendency and the halo effect. (see Chapter 3).
- Instead of using *Likert scales* (which assess agreement with statements using five categories ranging from strongly disagree to strongly agree) or adjectives that are left to individual interpretation (good, fair, poor), use actual descriptions of behaviors (referred to as *descriptive anchors* or behavioral anchors) that inform faculty what specific, observable behaviors should be demonstrated at each level. This is a way of standardizing the process. The following is an example of descriptive anchors for a question on documentation skills:

   Needs Improvement: Incomplete documentation, disorganized and/or misspelled words. Forgets, omits charting information.
   Competent: Appropriate documentation in medical record, clear, few misspelled words. Documents all necessary info.
   Exceeds Expectations: Neat and organized charting. Thoughtful plan that encompasses the whole patient. Includes details that are pertinent, and completely documents health info.

How can a resident have quite good scores on a set of standardized patient stations (OSCEs) and yet receive average global ratings by most of the faculty she has worked with?

Incongruent scores on assessments must be considered seriously. When two assessment methods disagree, it is possible that one is simply wrong. But it is also possible that both are wrong, or that both are right, but are assessing different (and yet still important) aspects of performance.

There may be several possible explanations for the pattern of assessment in the case example. The standardized patient exam may be flawed. Perhaps the standardized patients were not truly "standardized" or were not appropriately trained to assess the areas that are important to the scenario. There may have been, for example, only five OSCE stations, although evidence suggests that approximately 10-12 standardized patient stations, each lasting up to 25 minutes, are necessary for appropriate sampling.[11] On the other hand, the faculty ratings may be biased or lack validity for reasons already discussed. However, it is also possible that the differences between the OSCE scores and the faculty ratings reflect genuine differences in either the ability of the examinee to perform across varying contexts or the ability of the assessments to provide an overall summary of her performance.

Miller's pyramid, reproduced below, is often cited as a useful way to understand the concept of the progression of knowledge, skills and attitudes in medical education. The learner progresses from "knows" at the base to "does" at the top of the pyramid. The two steps between are "knows how", progressing to "shows". This model can be useful when thinking about types and locations of assessments. The American Board of Pediatrics In Training Exam (ITE) is a "knows" assessment, whereas an OSCE is at the per-

formance or "shows" level. Actual observation of clinical performance of a resident with a patient is at the action or "does" level. Performance may be different at each of these levels since different knowledge and/or skills are called into play at each of these levels. This also helps to explain the variability in the rotation evaluations when compared with the OSCE and how these may in turn vary from the ITE scores. Of course, the content of the exams may also be different.
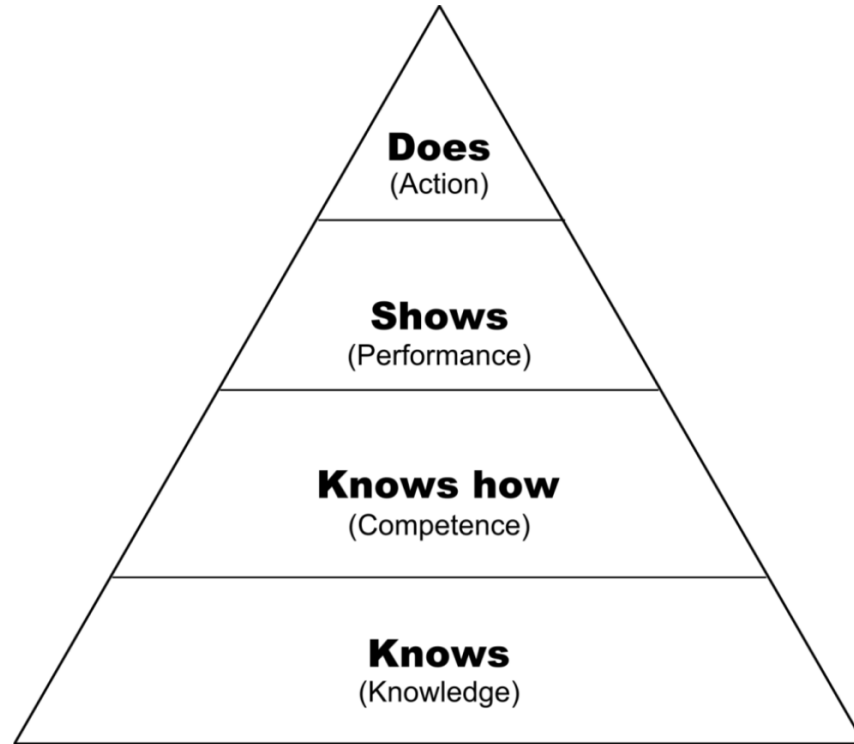
Miller's Pyramid. From Miller GE. 1990. The assesment of clinical skills/competence/ performance. Academic Medicine, 64(9):S63-7. Copyright © Association of American Medical Colleges. Reproduced with permission of Lippincott Williams & Wilkins.

It is important to note that validity evidence for assessment at the tip of the Pyramid can be difficult to establish. At this "Does" level, some control and standardization of the assessment setting is traded for unscripted *authenticity* of assessment.[5]

Should the resident be promoted? Does the committee have enough evidence to make a fair and meaningful inference about her progression to the next level?

It is important to weigh the validity evidence and try to minimize any threats to validity for this sort of decision. Discussions above have attempted to outline some maneuvers that will increase validity. *Triangulation*, or assessing and considering a resident from different perspectives (e.g., self-assessment, 360-degree evaluations, and patient surveys), can also help increase validity.

Assessments may be *formative* or *summative*. A formative assessment is designed to promote the development of the resident, usually through *feedback* on her performance. Formative assessments are commonly used in medical education as a tool for identifying a learner's strengths and weaknesses; weaknesses may be targeted for remedial instruction. A summative assessment, in contrast, is designed to provide a composite summary of a resident's performance that can be used to make decisions about the resident, such as assigning a grade for a course, or determining whether a resident should be promoted or kept back.

Using summative assessments to make decisions must be clearly thought out. For example, making an assessment that has important consequences, but lacks validity evidence would not be appropriate. Additionally, a program director should be familiar with the terms *norm-referenced* (comparing an individual's performance with that of a group of people performing similar tasks) and *criterion-referenced* (comparing an individual's performance to a predetermined standard).[5] In educational settings, criterion-referenced assessments can be used to try to determine (in absolute terms) whether the learner has reached an acceptable level of performance. For example, it is usually not helpful to compare one resident's ability to perform a complete and accurate history and physical examination with that of another resident. What you want to know is whether each resident can meet a standard that you have set for performing a complete and accurate history and physical examination. Accordingly, there must be pre-specified cut-off determinations for what scores would meet muster on a criterion-referenced test. In a *high-stakes* assessment like the American Board of Pediatrics ITE, the cut-points and validity of questions are determined through rigorous psychometric analysis. Conversely, a low-stakes assessment rarely warrants this level of rigor.

Maybe I should have each of my residents do multiple high-fidelity simulations for each domain I am assessing…but wait…that will cost too much and where will I find the time?

In the cycle of residency training we must consider the practicality and utility of the assessment system. A conceptual framework termed the *utility* model proposes that:

$$\text{Utility} = \text{reliability} \times \text{validity} \times \text{acceptability/practicality} \times \text{cost} \times \text{educational impact.}^{13}$$

This model highlights compromises between these parameters.[2] Depending on the purpose of the assessment, parameters may receive different weights. For example, high costs may be more tolerable if the assessment is one of a high-stakes nature. However, an assessment that is for feedback and is formative in nature, should be weighted heavier on the educational impact factor.[2] So program directors must consider practicality, cost and *feasibility*, in addition to reliability and validity, to make informed decisions about when, where, and how assessment of residents occurs within a program.

Lessons Learned

- Assessments are not valid or invalid, their degree of validity depends on the degree to which they yield evidence to support a judgment or decision.
- Resident assessment requires an understanding of validity and thoughtful planning to arrange a system that minimizes the common threats to validity.
- Reliability can be thought of as the consistency or reproducibility of an assessment.
- There are a number of practical strategies one can use to improve the evidence of validity of assessments such as faculty development, descriptive anchors on assessment forms, utilizing enough observations and enough trained raters, and triangulating many forms of assessment.
- Utility is a conceptual model for implementing assessment in programs. Cost, practicality, and educational impact need to be considered, along with validity and reliability.

References

1. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical Education* 2005; 39:309-317.
2. Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical Education* 2003; 37:830-837.
3. Streiner DL, Norman GT. *Health Measurement Scales: A Practical Guide to their Development and Use*. 3rd Edn. New York: Oxford University Press; 2003.

4. Walsh WB, Betz NE. *Tests and Assessments*. 4th Edn. New Jersey: Prentice Hall; 2001.
5. Downing SM, Yudowsky R. *Assessment in Health Professions Education*. New York: Routledge Group; 2009.
6. Downing SM. Reliability: on the reproducibility of assessment data. *Medical Education* 2004; 38:1006-1012.
7. Downing SM, Haladyna TM. *Validity threats: overcoming interference with proposed interpretations of assessment data*. Medical Education 2004; 38:327-333.
8. Downing SM. *Face validity of assessments: faith-based interpretations or evidence-based science?* Medical Education 2006.40:7-8.
9. Williams RG, Klamen DA, McGaghie WC. *Cognitive, social and environmental sources of bias in clinical performance ratings*. Teaching and Learning in Medicine 2003; 15:270-292.
10. Iramaneerat C, Yudkowsky R. Rater Errors in Clinical Skills Assessment of Medical Students. *Evaluation & the Health Professions* 2007; 30(3):266-283.
11. van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine* 1990; 2:58-76.
12. Holden JD. Hawthorne effects and research into professional practice. *Journal of Evaluation in Clinical Practice* 2001; 7(1):65-70.
13. van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education* 1996; 1:41-67.
14. Messick S. Validity. In: Linn RL, editor. *Educational Measurement*. MacMillan; 1989.

Annotated Bibliography

Downing SM. Reliability: on the reproducibility of assessment data. *Medical Education* 2004; 38:1006-1012.

An outstanding, concise description of methods of estimating reliability that are discussed in an intuitive and non-mathematically oriented manner. Easy to follow and has good examples. Downing reviews concepts of reliability and defines reliability in multiple ways and discusses how it relates to validity. He proposes that all assessment data must be reproducible in order to be meaningfully interpreted. He discusses how the type of consistency of assessment outcomes depends on the type of assessment. For example, written tests rely on internal consistency using estimation methods (test-retest design) while ratings of clinical performance require interrater consistency/agreement. The author concludes that reliability is a major source of validity evidence for assessments. Further, inconsistent assessment scores are difficult to interpret.

Downing SM. Validity: on the meaningful interpretation of assessment data. *Medical Education* 2003; 37:830-837.

This article discusses construct validity in medical education. It utilizes clear, applicable examples of validity evidence. The examples used are written and performance examination considerations, high-stakes and lower-stakes assessments. The author proposes and discusses that all assessments in medical education require evidence to be interpreted in a meaningful manner. All validity is "construct validity" and therefore requires multiple sources of evidence. He discusses the five types of construct validity evidence: content validity, response process, internal structure, relationship to other variables, and consequences. Thus, construct validity is the whole of validity, but has any number of facets. He also emphasizes that validity should be approached as a hypothesis. A key point from this article is that one must recognize that "assessments are not valid or invalid, but rather assessment scores have more or less validity evidence to support the proposed interpretations."

Downing SM. Face validity of assessments: faith-based interpretations or evidence-based science? *Medical Education* **2006; 40:7-8.**

This article discusses the term "face validity", reviews what it seems to mean, and discourages its use. It is defined as a vague property that makes an assessment appear to be valid, or look like it measured what it was intended to measure. The author says that this circular reasoning is a "pernicious fallacy".

Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* **2004; 38:327-333.**

The authors outline the factors that interfere with our ability to interpret assessment scores and/or ratings in the proposed or correct manner. They focus on two specific threats to validity: Construct under-representation (CU) and Construct-irrelevant variance (CIV). CU refers to undersampling of the content domain, i.e. using too few cases, items, or clinical observations to be able to adequately generalize. CIV refers to variables that systematically (not randomly) interfere with the ability to meaningfully interpret ratings or scores on assessments. The authors define these terms and give examples of these threats in medical education assessments (specifically ratings of clinical performance). They also discuss ways to minimize the validity threats in day to day residency assessment.

2. Assessment Methods

Patricia Hicks, MD

"Not everything that counts can be counted and not everything that can be counted counts."
    - Albert Einstein

Rationale

Program directors appreciate the importance of applying particular curricular and instructional methods to develop specific aspects of the learner's knowledge, skills and attitudes. For example, we know that a resident does not learn the complex set of behaviors and skills for effective teamwork by reading a book or listening to a lecture on how to develop teamwork skills. Rather, behaviors and attitudes are best shaped and developed through participation in role-play, interactive discussion groups, facilitated group sessions, simulation or through interaction that takes place within a real team where coaching provides course correction.

Just as one aligns instructional methods with the curricular content to be taught, it is important to align types of assessment tools with what one is trying to assess. In choosing an assessment instrument or tool, one must consider three facets of assessment: (a) what is the test content I am seeking and how should that content be proportionally represented and organized, [1] (b) what type of *assessment method* do I want to use,[2] and (c) in what *context* do I want to conduct the assessment?[3]

Test content choices are often approached using a process called *blueprinting*. Blueprinting is the process of defining the test content that one wants the assessment to sample. In general, the content should be chosen to represent those parts of the curriculum thought to be important; a test blueprint defines and precisely outlines the representative proportion of the test questions as they relate to the content areas of interest (knowledge or behaviors).[4]

The type of *assessment method* chosen should align with the nature of the knowledge, skills, or behaviors to be assessed. Assessing learner performance in giving bad news might be done using a trained and calibrated *standardized patient*, but using standardized patients to assess complex problem solving would not be effective because the ratings of the standardized patients do not capture the complexity of diagnostic reasoning and clinical decision-making. Such complex cognitive tasks are better assessed in well-constructed written examinations.

The *context*, or clinical setting or situation, in which the assessment is conducted should be carefully considered. We now know that learner performance achievement in one context does not inform us about that learner's performance in another setting [3]. For assessment to be meaningful, the data collected should represent performance close to the setting and conditions that would occur in the real world context.

Goals

1. Determine what you might want to assess and in what environment or context you might choose to conduct the assessment.
2. Choose the assessment method and tool that best align with the learning content that needs to be tested.
3. Understand the strengths and limitations of the various assessment tools you might consider using.

Case Example

A new curriculum has been developed in many areas of the hospital, focusing on various aspects of team functioning. The Neonatal Intensive Care Unit is working on interprofessional team member coordination in performance of the Neonatal Resuscitation Protocol; the inpatient ward teams are working on improving the team's

functioning on family centered rounds; and the Emergency Department's efforts are centered on the initial stabilization and management of the critically ill child.

You are asked to assess the teamwork skills of residents in your program. Your hospital has just opened a new simulation center and you are going to take advantage of this opportunity to develop a curriculum to teach resuscitation and initial stabilization of patients presenting in shock. You want to assess the skills of your learners. You decide that you will assign each a role at the start of the simulation scenario and intend to assess each learner's skills in performing the assigned role. You also decide to assess the ability of the team to function as a whole. In addition, you know that you will need to evaluate this new curriculum so that you can provide for ongoing programmatic improvement over time. Eventually, you would like to demonstrate that those who complete the simulation training and perform well on assessment also perform better in the actual patient care setting in the management of patients presenting to the Emergency Department in shock (compared with providers who have not yet received this training).

Points for Consideration

Now that we have looked at this case, let's consider some questions that many program directors face in the selection and interpretation of the data assessment tools yield.

How will you decide *what* you want to assess?

There are many facets to teamwork and the focus or area chosen to assess depends on the *stakeholder's* purpose of the assessment. For example, you may focus on (1) communication aspects of the team interactions, choosing to assess the use of a structured communication protocol, (2) whether this communication structure is used between team members and perhaps the quality or efficiency of its use in escalation of care for the critically ill patient, or (3) the learner's adherence to the clinical pathway or protocol chosen by your institution, identifying which steps the learner completed according to the criterion, whether extra steps (not indicated) were done, and whether interventions were performed correctly or not.

The tendency is to undertake assessment of everything possible so as to "test" all aspects of the learning. However, some clinical and educational settings lend themselves to some aspects of assessment more than others. Whatever aspects of learning you choose to assess, the data that result from such an assessment should be meaningful in regards to the real-world context in which the knowledge, skill, or attitude being assessed takes place. That is, the results of what you choose to assess should give you meaningful inference regarding the learner's abilities in the actual target setting.

What is the *purpose* of your assessment?

Assessment can provide *formative* feedback, giving learners an indication of their performance compared with desired outcomes. Ongoing formative feedback is key in facilitating developmental progression along a continuum of improvement. In the example of stabilization of a patient presenting in shock, such measures could be the ability of each team member to effectively and efficiently determine a component of the clinical status of the patient. One could focus on the learner whose role it was to determine respiratory status, another on the learner determining cardiac pump status and perfusion, and another on the learner determining mental/neurological status. Formative assessment might be conducted using a checklist and could be done in the simulated setting; immediate debriefing could be added so that subsequent exercises in the simulation center could build on the identified gaps.

If the assessment is to be used to determine which learners can be counted on to lead the team in the critical care room or carry a code pager, the assessment would be considered both summative and high-stakes. The purpose for this type of assessment might be to determine readiness to care for a patient presenting with shock, with minimal supervision. For a high-stakes assessment, the reliability should be very high (usually greater than 0.8) and the cut-point for such an assessment should be determined by an accepted standard setting method.

Critical evaluation of the curriculum delivered could be explored by determining if there were consistent errors by learner subjects during this standardized case assessment. Careful analysis of the aggregated learner assessments in the simulation center could lead the course developers to modify their instructional approach or the course content to address these areas of sub-optimal performance. Individual poor performance on assessment can differentiate the learner who did not learn; poor performance across the group of learners may indicate content or instructional gaps or ineffectiveness.

How will you *choose* an assessment method(s)?

Often, the selection of an assessment method is based on what is easy to use or create or is already available—even if that method is not the best one to gather evidence of performance for the identified behaviors. Ease of implementation and considerations of cost are critical, however, and thus the choice of method will be made within the limitations of resources of time, money, personnel and expertise[5].

You should select an assessment method based on an analysis of the type of behaviors for which you want to gather evidence. If you want to gather evidence about the use of a communication framework, you could video/audiorecord the interaction and then score the recording for the presence of identified elements. Similarly, you could score the behaviors using live observers, trained to "look and listen" for key elements of the communications. Other behaviors, such as procedures performed or other physical examination assessment maneuvers can also be observed live or by recording.

Written orders, documentation and other actions recorded in the electronic medical chart can be useful to determine the timing, sequence and comprehensiveness of carrying out the protocol or pathway. Combining evidence for plan-to-act and communications regarding assessment and plans with evidence of actual orders or written synthesis of patient status can be helpful in informing the assessor about the learner's ability to integrate tasks.

What *setting* will you use to perform assessment of teamwork skills?

Observations in the actual patient care setting may be used to assess learner skills or an artificial testing environment can be used to assess performance in a more planned and controlled setting. The advantage of directly observing teamwork behaviors in the actual patient care setting is that the setting is real and the behaviors witnessed are the actual behaviors in the patient care setting; no inferences need to be drawn. These conditions are often referred to as *authentic*.

A disadvantage of using the actual patient care setting for assessment is the lack of consistency in case-design and thus the wide variation in testing conditions, threatening the reliability and validity of the results of such assessments (see Chapter 1). With simulation, the setting can be controlled, raters calibrated, and reliability and validity of results can be quite high.[8] However, for simulated or standardized patient assessment to be useful, the design and methods used need to produce evidence of measurable behaviors that correlate closely with actions or outcomes demonstrated by a competent pediatrician in the actual clinical care setting.

What will you do with the *results* of (each) assessment?

Using the results of assessment to determine next steps for the learner requires high validity evidence and clear communication to the learner that the assessment is going to be used for determining improvement opportunities or for some other consequence. The very act of associating an assessment with a high-stakes outcome can pose a threat to the validity evidence of that assessment. Knowing that one is being assessed often influences performance; only the most expert of performers is accustomed to perform optimally in high-stakes settings.

Formative assessment of learner performance can be used to frame a debriefing discussion, where the individual or group can explore what was done well in addition to identifying areas for improvement. Group discussion

and individual reflection can combine to result in highly effective learning, adding further value to this type of assessment. Coaching, debriefing and reflection use assessment to drive and direct learning.

It is not clear that performance in one clinical context can be used to give meaningful inference to learner performance in another setting. If the setting used is not the actual patient care setting, it must be carefully constructed so that evidence gathered measures what you intended to measure.

<u>What limitations or challenges will you recognize in your choice of assessment method?</u>

A *checklist* completed by a rater is often the assessment method used for scoring performance of behaviors. Challenges of such an assessment method include: checklist item construction, rater training and rater *calibration*. Identifying the types of items and then describing those items specifically can be difficult. Rater training is critical so that each rater scores observed behaviors correctly and consistently. The assessor is often challenged with developing faculty so that the tools designed are used in a consistent, precise and reliable manner so that the data generated have high validity evidence. Engaging the faculty in assessment methods will often result in increased awareness of the curricular design and instructional methods. Consideration of faculty culture, including responsiveness to change is an important step.[6]

Self-assessment can be used for program feedback regarding the learner's perception of their response to the curriculum or the assessment (whether they found it fun, safe, well organized, thought-provoking, and such) but self-assessment is rarely useful as a method for accurately determining learner performance. However, the process of self-assessment, if done with a mentor, can help the learner gain insight and models reflective practice. See Chapter 4 for further detail.

Assessment in simulated, as opposed to real-life, clinical settings can be limited by the *Hawthorne effect* (the tendency of research subjects to behave atypically as a result of their awareness of being studied, as opposed to behaviors occurring as a result of the actual treatment that has occurred), and by definition, assesses what the learner is capable of rather than what they actually do in the real-life setting.

Patient clinical outcome measures are the most useful measures of the results of physician ability and assessing outcomes is superior to testing of sub-components of larger, integrated tasks. Thus, a limitation of assessing components of teamwork, such as communication, sequencing of tasks, etc., is such assessments fall short of measuring the overall outcome of a clinical task such as stabilizing the critically ill patient. Assessing relevant subcomponents after sub-optimal global performance may be more useful. The overall task is often more than the sum of the parts, but assessing the parts can be useful in determining what is missing in the overall performance.

Kern et al.'s seminal book on curriculum design[2] also reviews the uses, strengths, and limitations of a variety of assessment methods. The table on the next page summarizes the recommended uses and strengths. As discussed above, each tool also has limitations.

Lessons Learned

- The selection of assessment tools should align with the type of performance to be assessed. For example, direct observations of behavior are required to assess *presence of desired* behaviors and to assess *absence of undesirable behaviors.*[7]
- Choice of assessment methods should be based on what provides the best evidence of achievement or performance and not what is easiest to measure.
- In addition to reliability and validity, cost, practicality and impact are important factors to consider when choosing an assessment tool.

- Results of assessment drive improvement in curriculum development and instructional design; development of assessment methods and curriculum are iterative processes, each informing the development of the other.
- Assessment in a simulation center allows for a standardized testing environment where the prompts and conditions are well controlled, but assesses capability rather than typical performance.[8]
- Assessment results from one setting may not give meaningful evidence of outcomes in other settings.[9]
- The use of assessment tools and interpretation of data from those assessment tools requires faculty development if results of assessment are to be reliable, valid and meaningful.[10]

	Best used to assess				Strengths				
	Attitudes	Knowledge	Skills	Performance	Low cost	Appropriate for formative assessment	Accepted as summative assessment	Objective	Learner-centered
Rating forms	Y	Y	Y	Y	Y	Y	Y	N	N
Self-assessment	Y	Y	Y	Y	Y	Y	N	N	Y
Essays/journals	Y	Y	N	N	Y	Y	N	N	Y
Written or computer-based constructed response tests	N	Y	N	N	Y	N	Y	Y	N
Oral exams	Y	Y	N	N	N	Y	Y	N	Y
Direct observation including OSCEs	N	N	Y	Y	N	Y	Y	N	N

Recommended uses and strengths of common assessment methods.

References

1. Raymnod M, Neustel S. Determining the content of credentialing examinations. In: Downing SM, Halyadyna TM, ed. *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
2. Kern DE, Thomas PA, Howard DM, Bass EB. *Curriculum Development for Medical Education: A Six Step Approach*. Baltimore and London: Johns Hopkins University Press; 1998.
3. Norman G. How specific is case specificity? *Medical Education*. 2006; 40:618-623.
4. Linn RL. The standards for educational and psychological testing: Guidance in test development. In: Downing SM, Haladyna TM, ed. *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.
5. van der Vleuten CPM. The assessment of professional competence: developments, research, and practical implications. *Advances in Health Sciences Education*. 1996; 1:41-67.
6. Bland CJ, Starnaman S, Wersal L, Moorehead-Rosenberg L, Zonia S, Henry R. Curricular Change in Medical Schools: How to Succeed. *Academic Medicine*. 2000; 75(6):575-594.
7. van der Vleuten CPM. Validity of final examinations in undergraduate medical training. *British Medical Journal*. 2000; 321:1217-1219.
8. Tuttle R, Cohen M, Augustine A, Novotny D, Delgado E, Dongilli T, Lutz J, DeVita M. Utilizing simulation technology for competency skills assessment and a comparison of traditional methods of training to simulation-based training. *Respiratory Care*. 2007; 52(3):263-.
9. Lawson D. Applying generalizability theory to high-stakes objective structured clinical examinations in a naturalistic environment. *Journal of Manipulative Physiological Therapy*. 2006; 29:463-467.

10. Downing SM, Yudkowsky R. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009.

Annotated Bibliography

Downing SM, Yudkowsky R. *Assessment in Health Professions Education***. New York, NY: Routledge; 2009.**

This well organized and very readable book is the next step for any program director or medical educator wanting to know more about assessment in health professions. Drs. Downing and Yudkowsky have done a superb job of explaining complex concepts in easily understandable terms, with examples that any program director can relate to and apply. The book is written for every learner level from the novice in assessment to those who are expert and want to explore further resources offered in the rich number and quality of references cited. If one is feeling uncertain or unprepared to learn more about assessment, the response should be to get this book and start the exploration and discovery of assessment!

Kern DE, Thomas PA, Howard DM, Bass EB. *Curriculum Development for Medical Education: A Six Step Approach***. Baltimore and London: Johns Hopkins University Press; 1998.**

This 180 page book translates theory to application in a step-wise fashion. The title may be a bit misleading in that the book really describes the full cycle from problem identification and general needs assessment to specific goals and measurable objectives to educational strategies, implementation evaluation and feedback. While evaluation of the curriculum is emphasized, individual learner assessment is the focus for some aspects of curriculum evaluation and thus this text may be useful to those wanting to learn more about assessment. Complex interactions between instructional methods, learning content and outcome measures are explained in simple terms and with many practical examples. The references are classic, but a bit dated in this 1998 publication. This is a must read for any program director!

3. Faculty Development

Julia McMillan, MD

"To be effective teachers, faculty require diverse skills such as creating a facilitative learning environment, observing and assessing learners, providing feedback, teaching in small groups, lecturing, mentoring, and developing and evaluating curricula. Such skills can be taught effectively, but most faculty have not received formal training in them."[1]

Rationale

The word "doctor" means "teacher" in Latin, yet instruction in teaching skill and assessment of learners are not overt components of the medical school curriculum. Residents and fellows are expected to teach, and increasingly residency and fellowship programs have included instruction in adult learning principles. Teaching and assessment of students, residents, and fellows are essential faculty responsibilities, yet most faculty members face these challenges with little preparation, guidance, or understanding of expected results. Most faculty members have benefited from teachers they felt were effective and engaging, but they are unlikely to have considered the components of communication and behavior that contributed to their effectiveness.

As regulatory bodies have enhanced requirements for specified curricula and for assessment of competence, medical schools and residency programs have responded by increasing expectations for faculty time and effort devoted to education. Without prior training and with little understanding of the elements of effective medical education, faculty members may resist involvement with students and residents; or they may continue teaching as they always have, hoping that they are mimicking the behaviors of effective teachers they remember. When challenged to provide specific verbal feedback and competency-based assessment, they may feel intimidated and even resentful at the suggestion that their efforts are no longer considered adequate. It is within this context that medical educators have dissected the clinical education process, developed a variety of programs intended to enhance the educational efforts of their colleagues, and attempted to evaluate the effectiveness of those programs. On an institutional level, it is important that faculty members who contribute to efforts that enhance the effectiveness of their colleagues as teachers receive support and recognition that is equivalent to that provided for faculty in other academic areas.

Goals

1. Understand the various opportunities and settings in which faculty serve as teachers and provide assessment for medical students and residents.
2. Learn principles for effective faculty development.
3. Gain appreciation for the value of engaging faculty members in activities that enhance their effectiveness as teachers and assessors.
4. Identify anticipated outcomes for effective faculty development programs.

Case Example

One of your most experienced clinical faculty members comes to tell you that he feels he will no longer be able to serve as precepting attending on the general inpatient service. In the past his responsibilities included clinical supervision, teaching, and assessment of resident performance. In recent years, however, he has been frustrated and overwhelmed by the requirement that he assess residents with regard to their achievement of specified goals in the six core competency areas defined by the Accreditation Council for

Graduate Medical Education (ACGME) and provide ongoing feedback in relation to each resident's achievement of those goals. He feels inadequately prepared for these new responsibilities, and he doesn't understand why past education methods are no longer adequate. You confer with your department chair, and together you decide that this faculty member is too valuable as a teacher and role model to allow him to give up without first offering support that will help him continue in his role while adapting to the more rigorous medical education structure.

Points for Consideration

What can you tell this faculty member to reassure him that his knowledge and experience is critical to helping make individual competency assessment relevant to everyday clinical activities?

Whatever the forum in which faculty development activities occur, faculty participants should be reminded that their own experiences and clinical instincts are valuable resources for both teaching and assessment. The terminology of the ACGME's six *core competencies* may seem artificial, and separating them as distinct aspects of clinical care creates a compartmentalization that is foreign to the intuitive judgment of faculty members accustomed to assessing trainee skills as activities that may involve many specific competencies simultaneously (see Part II). It is important to acknowledge that the effective integration of those competencies is, in fact, the goal of training, even though assessment may require consideration of individual competencies. The faculty member who is informed about the meaning of the core competencies and is actively engaged in teaching in a clinical setting will recognize each of those competencies in the everyday activities of the residents with whom they work. An important goal for faculty development activities is to achieve agreement among faculty regarding appropriate standards to be used for learner assessment at various stages of development as pediatricians.

How can this faculty member fit best into the clinical educational activities of the program?

Clinical educators have opportunities to teach and assess residents in a variety of settings, including all of the following:

- Lectures
- Facilitation of small groups sessions
- Simulated exercises and teaching procedures
- Bedside teaching
- Inpatient and outpatient precepting
- Mentoring
- Curriculum development

It is important to recognize that particular faculty members may be more comfortable and effective in some of these settings than in others. Development of skilled and confident faculty educators in each of these settings requires somewhat distinct faculty guidance, but the principles for success are similar: (1) determine the gaps between the current knowledge, skills and attitudes of faculty and the program goals (*needs assessment*), (2) set goals and expectations, (3) convey information, observe performance, and encourage self-education, (4) provide interim feedback that is both corrective and reinforcing, and (5) deliver a final assessment that includes suggestions for next developmental steps.

<u>You've identified the forum in which this faculty member's skills are most effective. What's the best way to ensure that he is as effective as possible and that he develops the confidence needed to feel successful as a clinical educator?</u>

There are a variety of formats for faculty development. Each has limitations and strengths, and all could be effective, depending on the audience, the gaps to be bridged, and the goals to be achieved.[1] Lectures and discussions at faculty meetings are an efficient means of informing faculty members about broad curricular changes, reminding them about overall responsibilities of the teaching faculty, including responsibility for careful assessment of learners, and providing feedback regarding the successes or shortcomings of particular aspects of the program. These discussions, however, do not specifically address either the setting in which each faculty member will teach or the strengths and areas for improvement for individual faculty members. Large group meetings of faculty members may serve as a first step in the process of developing more intense or individualized interventions to improve teaching methods and standards for assessment. If enough well-informed faculty leaders are available, faculty meetings can even be an opportunity for small group break-out discussions. For the faculty member in the vignette and others like him, it is important that lectures and discussions before large groups of faculty members do not add to the frustration he is already feeling. This can be avoided by describing the support that will be available to implement proposed changes in responsibilities and by providing descriptions of the incremental steps that will be taken.

Seminars and workshops provide an opportunity for faculty members to problem-solve through role play exercises, small group discussions, and sharing of personal experiences. These sessions should target a particular element of faculty development, such as organizing small group learning opportunities, providing timely and effective feedback, finding agreement on standards for assessment of individual competencies, or conducting family-centered rounds. The elements that are important in making these sessions productive include (1) a leader who has clear goals in mind for the participants and can set the stage for the discussion, and (2) willingness on the part of participants to share questions and discuss both their successes and their challenges. The faculty member in the case example is more likely to be willing (and perhaps even enthusiastic) to participate in these exercises if he understands that they will be relevant to the clinical teaching format with which he is involved and will include other faculty members engaged in similar activities.

Peer teaching and assessment, including team teaching, allows mutual observation and feedback for teachers. This approach requires a positive learning environment in which faculty members are mutually supportive and communicate clearly, and it is a time and effort-intensive process. Using a formal peer assessment tool, Beckman, et al, compared peer evaluation by three clinician faculty members of one of their colleagues during rounds on an internal medicine inpatient service.[2] The description of the process engaged in by the participants in this study may be more instructive than its conclusions. Peer evaluators differed regarding their understanding of the goals for teaching in this setting. In addition, evaluators' standards for quality in the various aspects of the teaching session were not the same, and some evaluators highlighted aspects of the session that others did not notice. The process, however, allowed educators to discuss and reflect on their own criteria and goals as teachers as well as the accuracy of their assessment of their colleagues. Another approach to peer guidance in teaching involves pairing of faculty members in co-teaching teams, allowing each member of the team to learn from the other's methods.[3] In the best of circumstances, co-teaching also allows mutual opportunities for constructive and supportive feedback from colleagues. This type of activity would also allow two or more faculty members to compare their assessments of learners.

Periodic meetings of faculty members engaged in similar teaching activities, such as attending on the inpatient ward, continuity clinic supervision, or precepting in the emergency department or intensive care unit, allows teachers and evaluators to come to agreement about methods of teaching and standards for

assessment.[4] These groups can provide support and collaboration for educational efforts within their own rotation. Involved faculty can develop protocols for direct observation of residents, opportunities for residents to investigate and report on patient care practices (including journal club), and standards for providing feedback. Calibration of faculty rating on evaluation forms is also possible when faculty come together to agree on performance standards.

Longitudinal courses allow time for engaged faculty members who are leaders in curriculum development or educational projects to develop the knowledge and skill needed for success.[5] When such courses are effective they allow consideration of a variety of teaching methods in an array of possible settings and focus attention on the needs of the specific learners with whom the faculty member is involved. Such longitudinal courses require significant faculty time that can be scheduled to coincide with the course requirements. For faculty members who enjoy clinical teaching, such as the individual in the case example, such courses can clarify the language and the challenges brought about by the current, more prescriptive approach to teaching and assessing residents.

<u>How might you assist this faculty member in his fair and accurate assessment of residents? How can you be certain that the same standards are being used by all evaluators?</u>

Bias and inconsistency are introduced into the assessment process in a variety of ways: (1) the observer may have only a limited time to witness the resident's clinical performance; (2) the assessment may be influenced positively or negatively by the observer's prior experiences with the resident; (3) the assessment may be completed days or weeks after the observation, making it difficult for the observer to recall performance accurately; and (4) observers may have differing expectations with regard to a resident's knowledge and clinical development. Assessment that is completely unbiased, accurate, and standard across multiple observers is probably not an achievable goal in the context of residency training; however, there are several best practices:

- Multiple observations by multiple observers in several clinical settings enhance the accuracy of conclusions regarding clinical competence. Some research suggests that assessment by at least 7 to 11 observers leads to a reproducible assessment.
- Assessment of specific performance attributes (desirable or undesirable) should be provided as soon as possible after the observation and not be delayed until written comments are requested.
- Recording of comments at the time of observations during a rotation (diary keeping) allows the observer to provide a more complete and accurate assessment when the clinical performance evaluation is requested.
- At the beginning of a rotation, the rating form and the frame of reference for expected performance should be explained to faculty members who are expected to assess residents.
- Feedback should be provided to the faculty assessors to give them an understanding of their severity or leniency in assessment relative to other faculty members.[6]

<u>What is known about the impact of faculty development activities, such as those described above, on the quality of teaching and assessment of learners, or on patient care?</u>

Studies of the outcome of faculty development are most consistent in demonstrating improvement in faculty attitudes toward teaching and in self-awareness of faculty strengths and limitations.[7-8] Participants report enhanced knowledge of educational principles and improved teaching skills, both self-perceived and as reported by their learners. Little is known about the impact of improved teaching and assessment

on the outcome of learner knowledge and skill development.[9] Similarly, little is known about the impact of faculty development efforts on organizational structure and culture or on patient care.

Lessons Learned

- Changes in curriculum and assessment of learners require meaningful faculty development programs and depend on leadership and institutional/departmental support.[9] Participation of faculty members in efforts to enhance teaching effectiveness and achieve consistency in assessment requires consideration of practical issues, such as time constraints, competing commitments, and availability of faculty development leaders.
- Although effectiveness of faculty development programs can be measured by assessment of learners and through its impact on the culture in which it takes place, additional studies are needed to verify the outcomes.
- Medical education is a scholarly discipline, and experimentation and research are required to assess current practices, to enhance the skills of our faculty, and to link educational programs with improvements in patient care.

References

1. Clark, JM, Houston TK, Kolodner K, Branch WT, Levine RB, Kern DE. Teaching the teachers: National survey of faculty development in departments of medicine of U.S. teaching hospitals. *Journal of General Internal Medicine* 2004; 19:205-214.
2. Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Medical Teacher* 2003; 25:131-135.
3. Orlander JD, Gupta M, Fincke BG, Manning ME, Hershman W. Co-teaching: a faculty development strategy. *Medical Education* 2000; 34:257-265.
4. Hemmer PA, Pangaro L. Using formal evaluation sessions for case-based faculty development during clinical clerkships. *Academic Medicine* 2000; 12:1216-1221.
5. Cole KA, Barker LR, Kolodner K, Williamson P, Wright SM, Kern DE. Faculty development in teaching skills: an intensive longitudinal model. *Academic Medicine* 2004; 79:469-480.
6. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings *Teaching and Learning in Medicine* 2009; 15:270-292.
7. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Medical Teacher* 2006; 28:497-526.
8. Berbano EP, Browning R, Pangaro L, Jackson JL. The impact of the Stanford Faculty Development Program on ambulatory teaching behavior. *Journal of General Internal Medicine* 2006; 21:430-434.
9. Bland CJ, Starnaman S, Wersal L, Mooread-Rosenberg L, Zonia S, Henry R. Curricular change in medical schools: how to succeed. *Academic Medicine* 2000; 75:575-594.

Annotated Bibliography

Clark, JM, Houston TK, Kolodner K, Branch WT, Levine RB, Kern DE. Teaching the teachers: National survey of faculty development in departments of medicine of U.S. teaching hospitals. *Journal of General Internal Medicine* 2004; 19:205-214.

The authors surveyed U.S. hospitals with internal medicine residency training programs to determine the presence or absence, and the nature of faculty development activities. Almost 72% of the 389 existing programs responded to the survey. Thirty-nine percent (N=108) of the programs reported ongoing faculty

development activities, with university hospitals more likely to have such programs. Small group discussions were the most prevalent teaching method, but role plays, observation with feedback, teaching projects, standardized patients, and simulated learners were also used. Most activities were offered as half-day workshops, but some hospitals used ongoing courses of varying duration. More intense programs were more likely to be supported by offering salary support or protected time to faculty development instructors. CME credit was offered by 47% of hospitals. Evaluation of programs was most often carried out using evaluation forms completed by participants. Responses to open-ended questions on the survey indicated that expertise in teaching skills, time, funding, and infrastructure were all important in determining the availability of faculty development activities. In addition, the value placed on teaching by the institution or department was seen as an important influence on faculty development activities.

Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Medical Teacher* 2003; 25:131-135.

The authors developed an assessment instrument, derived from the Stanford Faculty Development Program framework (SFDP-26) for use in peer assessment of faculty clinical teaching. The new instrument, the Mayo Teaching Evaluation Form (MTEF-28) is a 28-item instrument with 5 options per item. The MTEF-28 was validated by applying it to the assessment of 10 faculty attending on the internal medicine service. Three faculty members observed teaching rounds led by each of the 10 faculty members for one morning and completed the form within one hour. Seven general categories were assessed for validity: learning climate, control of session, communication of goals, understanding and retention, evaluation, feedback, and self-directed learning. The highest alpha scores, indicating greater reliability compared to the SFDP-26, were in the categories of self-directed learning, learning climate, communication of goals, and evaluation. The authors conclude that the MTEF reliably assesses teaching behaviors within the Stanford educational framework, that there was significant agreement across the Mayo evaluators, and that use of the MTEF provided peer evaluators with valuable insights regarding their understanding of the principles of effective teaching.

Hemmer PA, Pangaro L. Using formal evaluation sessions for case-based faculty development during clinical clerkships. *Academic Medicine* 2000; 12:1216-1221.

This article describes a faculty development technique that promotes an opportunity for discussion of medical student teaching and evaluation. The clerkship directors at the seven inpatient teaching sites for the internal medicine clerkship of the Uniformed Services University, along with residents and clerkship preceptors, meet monthly to discuss the goals for the clerkship, illustrate teaching methods, and review the evaluation of the students on the rotation. The authors describe these meetings as opportunities for discussion of an appropriate learning climate, leadership styles, and communication of goals for student learning and behavior. As specific students are discussed there is an opportunity to discuss evaluation methods and principles of providing feedback to learners. General issues of medical education are discussed, including evidence from education research. The authors argue that this type of session provides both a formal, planned, and longitudinal format for student evaluation and feedback, as well as real-time faculty development.

Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings *Teaching and Learning in Medicine* 2009; 15:270-292.

This comprehensive review describes the cognitive, social, and environmental factors that contribute unwanted sources of variation in scores in clinical performance assessments. The authors describe the various contexts in which performance assessments are carried out and review the available evidence that bias is intrinsic to that process. They describe the available evidence for mechanisms that reduce bias, and they

conclude by extrapolating sixteen recommended strategies for improving clinical practice assessments based on studies in both medical and non-medical contexts.

Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Medical Teacher* **2006; 28:497-526.**

This article was an effort of the international group, The Best Evidence Medical Education (BEME) Collaboration, to synthesize the existing evidence to answer the question, "What are the effects of faculty development interventions on the knowledge, attitudes and skills of teachers in medical education, and on the institutions in which they work?" The authors reviewed 2777 abstracts and ultimately found 53 articles that addressed improvement in clinical and basic science teaching and included outcome data that did more than poll participants regarding their satisfaction. The majority of the articles reviewed were from the United States, but some were from Canada, Egypt, Israel, Switzerland, Malta, Nigeria, the United Kingdom, and South Africa. Instructional methods, duration of programs, and type of program (seminar, workshop, short course, etc) are summarized. Outcomes for the various types of interventions are categorized according to Kirkpatrick's model for evaluating educational outcomes: Level 1, *Reaction* (impact on the participants' view of the learning experience); Level 2A and 2B, *Learning* (change in attitudes or perceptions [A] or acquisition of concepts, procedures, principles, or skills [B]) by participants; Level 3, *Behavior* (willingness of learners to apply new knowledge and skills); and 4A and 4B, *Results* (wider changes in the organization [A] or improvement in learning or performance [B]). The eight studies that achieve the highest rating according to this stratification are summarized. The authors conclude that, though there are multiple methodological limitations to the articles reviewed, satisfaction with faculty development programs is high. In general, participants report a greater awareness of personal strengths and limitations, enhanced motivation and enthusiasm for teaching, increase in knowledge of educational concepts and principles, and self-perceived changes in teaching behavior. Very few studies investigated the impact of faculty development on organizational practice or on student learning. In reviewing their findings, as well as previous literature, the authors list several components of faculty development that contribute to its effectiveness. They include opportunities to practice what has been learned and receiving feedback, collegial support for promotion and maintaining change, adherence to principles of adult learning, and use of multiple instructional methods. The authors identified several limitations to the existing faculty development literature and implications for future research, including a need for more rigorous study design, development of reliable and valid measures of change, assessment of maintenance of change over time, effective comparison of faculty development strategies, and a need to assess the impact of these programs on the institution or organization.

4. Self-assessment

Suzanne K. Woods, MD

"Nowhere can man find a quieter or more untroubled retreat than in his own soul." - Marcus Aurelius

Rationale

Physicians are charged with the responsibility to determine their individual learning needs and identify resources to aid in their personal education. Self-assessment is the ability of an individual to evaluate his performance, skills, and personal and professional qualities. Studies suggest that physicians have a limited capability to perform accurate, consistent self-assessment. To a self-regulating profession, this presents a challenge and offers opportunities for improvement. In clinical practice, determining one's abilities to safely care for patients is important in the delivery of high quality, safe patient care. One must be cognizant of one's own current and developing clinical competence in order to identify limitations or gaps in knowledge, skills, or attitudes. Despite some caveats, self-assessment is an important skill and may provide a critical stimulus in daily physician activity to reflect upon and review one's clinical practice. For this evaluation to be most meaningful and valid it is best to compare it to objective criteria.[1] Self-assessment can help to identify areas needing further education. It is important for physicians at all levels to understand their limits at the point of care and allow for reflection in practice. Medical educators play a critical role in fostering the development of self-assessment skills of learners.

Goals

1. Identify the purpose of *self-assessment* and recognize its importance in providing safe and quality patient care.
2. Describe the concepts of *reflection-in-action* and *reflection-on-action* and incorporation of these strategies into the context of self-assessment.
3. Understand the limitations of self-assessment.
4. Develop methods to make self-assessment meaningful and practical.

Case Example

As the program director, you are very interested in improving your residents' skills in giving bad news. You have been using survey feedback from nurses, parents, patients, and social workers on the team to assess these skills. Some of these survey tools measure physician empathy, others focus on communication skills, and other tools address cultural sensitivity. You wonder about the value of having residents self-assess their effectiveness in giving bad news. You consider several questions: (1) what would self-assessment add to a resident's insight into his performance? (2) will the comparison of self-assessed performance with ratings from parents, patients, nurses and social workers increase the resident's understanding of how others perceive his behaviors? (3) does self-assessment prior to giving bad news to a patient change the resident's performance in giving bad news? (4) what should be done if self-assessments are discrepant from assessments by peers, patients, or co-workers?

Points for Consideration

What is self-assessment and what is its purpose?

Self-assessment has been defined in many ways. These include:

- "the involvement of learners in judging whether or not learner-identified standards have been met"[2]
- "a process of personal reflection based on an unguided review of practice and experience for the purposes of making judgments regarding one's own current level of knowledge, skills, and understanding as a prequel to self-directed learning activities that will improve overall performance and thereby maintain competence"[3]
- "the ability to draw general conclusions about one's skills or knowledge in specific domains"[2]
- "self-rating" and "self-audit" activity[3]
- "a broad process of self-directed evaluation that is initiated and driven by the individual and is used for ongoing improvement"[4]

Although there are many definitions of self-assessment, similar themes emerge as components of the self-assessment process. Simply said it is "a personal evaluation of one's professional attributes and abilities against perceived norms".[4] External feedback and practice over time allow for the accuracy and validity of self-assessment to develop. Gordon describes valid self-assessment as "judging one's performance against appropriate criteria", and accurate self-assessment as "gaining reasonable concurrence between self-claimed and other, validated, measures of performance".[5] The purpose of self-assessment is to improve the perception of learning needs, promote change in learning activity, improve clinical practice, and improve patient outcomes.[4]

Self-assessment plays a role in each of the ACGME core competencies. By using self-assessment tools, physicians can improve insight into their performance. In the case example, residents may use several tools to evaluate past experiences with communication of bad news to patients. One tool is writing a reflection piece on one or more specific similar patient encounters that previously occurred. This should be paired with feedback from others who participated in those experiences including colleagues, faculty, patients, and caregivers. Other tools include reviews of videos of past performances, and participation in learning activities such as lectures or workshops about the topic of interest. Completion of a checklist or survey following an event may encourage self-assessment. All of these methods can aid the resident to gain insight into his performance. Improvements can then be made to enhance the experience of delivering bad news.

Why is self-assessment important?

The accurate identification and understanding of one's strengths is important for a physician to practice with confidence.[2] In addition, knowing one's strengths allows an individual to set appropriate personal learning goals and strive for improvement. If one can identify areas of relative weakness then this awareness allows for a self-imposed limit of one's competence in a particular patient encounter. Subsequently, one can seek out resources and references to aid with a particular task. This ability to identify one's weakness is also believed to "serve the function of helping the professional set appropriate learning goals".[6]

In combination, the ability to identify one's strengths and weaknesses for one's self can help develop "situation-specific" self-awareness.[3] Eva and Regehr write that self-assessment "generates a balance of confidence and caution, of persistence and flexibility, of experimentation and safety, and of independence and collaboration".[2]

How and when should self-assessment take place?

Self-assessment is best done using reflection, both in practice and on practice, and in comparison to peers and accepted standards of practice. *Reflective practice* is an important tool for physicians. It is the ability

to critically examine one's own reasoning and decisions. *Reflection-on-action* is the reflection that occurs following an event or experience and incorporates one's current knowledge of a situation or problem and also addresses how the situation could have been handled differently.[7] This can be viewed as a summative evaluation of one's actions. It is largely based on past experience, knowledge, and beliefs.

In contrast, *reflection-in-action* is "a task-bound reflective process in which we continue to act but re-shape our action online through explicit cognition".[2,7] This may occur unconsciously while one is in-volved in a particular task. Development of this skill assists with the provision of safe and effective patient care. Physicians need to be constantly aware of specific skills and knowledge that are needed in order to make safe patient care decisions in the moment. In the case example, knowledge and consideration of self-assessment items prior to giving bad news to a patient may enhance the resident's performance in communication of such news. Also, reflection-in-action is important and includes one noting the use of body language, tone, expression and emotion shown, and the response of the recipient to the bad news. One needs to monitor self and have situational awareness, reassessing the discussion frequently to adjust the presentation of the bad news and incorporate feedback and information from others. This "in the moment" capacity to identify what questions to answer and how to do so, as well as knowing what questions to defer until more information is gathered is fundamental in self-assessment.[8] This reflection-in-action is a dynamic and ongoing monitoring process. Research on reflection in practice (e.g. knowing one's limits at the point of contact) is promising, in that people appear to be more accurate in reflection-in-action than reflection-on-action. It has also been suggested that reflective practice can improve clinical judgment, minimize diagnostic errors, and aid in the development of medical proficiency.[9] Self-assessment should include measures of knowledge and reasoning but also other domains (behaviors, skills, processes of care, outcomes).[10] Residency training is a critical time for this understanding to occur.

Eva and Regehr describe two concepts which influence self-assessment. One is *self-concept* which is "a relatively sweeping cognitive appraisal of oneself".[2] This is formed by external feedback and introspec-tion. *Self-efficacy* is another influence and is defined as "a context-specific assessment of competence to perform a specific task" and "an individual's judgment of her capabilities to achieve a given goal".[2]

A critical element in the development of self-assessment skills is the incorporation of reliable and valid external sources of feedback. Outside feedback can be of particular help when providing information about communication skills, professionalism, and interpersonal skills and behaviors. Peer and other exter-nal feedback including that obtained from nurses, parents, social workers, and patients should be taken into consideration to understand how others perceive the resident's work, behaviors, and effectiveness. This feedback is best utilized when given over an extended period of time.

What are challenges of self-assessment?

Despite many reports citing the value of self-assessment, most studies reveal that an individual's ability to perform self-assessment is poor. Physicians that perform least well by external assessment demonstrate the greatest difficulties with self-assessment and are most likely to overestimate their abilities. These learners lack the knowledge of their own knowledge (*metacognition*) to accurately assess their own per-formance. In contrast to poor performers, high performers tend to underestimate their abilities more than external reviewers, in part because they assume that others also perform at their high level.[3,11] This is known as the "false-consensus effect". The inaccuracy of self-assessment for both high and low perform-ers results in *miscalibration*. However, a residency program director can help reduce miscalibration. The program director can highlight successes and problems of an individual resident and share multi-rater ex-ternal evaluation in a way that is instructive and constructive to trainees in developing their own self-assessment.

Although discrepancies between self-assessment and assessment by peers, patients, and co-workers represent a significant challenge, cognizance of these discrepancies and conscious development of self-assessment skills during residency training may aid learners to improve their ability to self-assess before they enter unsupervised practice where external feedback is more difficult to obtain.

How can program directors use self-assessment in residency training?

Self-assessment can be introduced in residency training in a variety of ways. These include having residents write a self-reflection piece about a difficult patient encounter or memorable experience. Also, the use of *multi-source (360-degree) assessments* aids in identifying one's strengths and weakness and can offer useful information about communication and interpersonal skills, professionalism, and behaviors. Another tool that can be implemented is the individualized learning plan (ILP). An ILP requires the learner to assess one's strengths, identify areas for improvement and learning needs, create learning goals, identify strategies to accomplish these goals, and document progress in the achievement of these goals. The components of the ILP become more tangible when reviewed with an advisor, mentor, colleague, or program director. Faculty monitoring of the progress of ILP completion may help in developing a resident's ability to proactively develop effective, self-directed lifelong learning which is critically important as a physician.[12]

What else do we need to know about self-assessment?

In order for self-assessment to play a significant role in continuing professional development, improvements need to occur in self-assessment methods. An increase in the awareness by practicing physicians of the need to seek out feedback from peers and other external resources will be critical in aiding their ability to more completely assess their performance and identify areas for improvement. External validation should be routinely used to aid in the process of self-assessment. This includes feedback from patients, peers, and supervisors, as well as comparison to consensus-based performance standards that are available with the goal of improving quality of care. Also, studies will need to be conducted of the attitudes of physicians toward self-assessment, especially in the changing era of maintenance of certification and the need to demonstrate continued competence in patient care and performance to meet quality measures.

Lessons Learned

- Physician professional development is often dependent on the ability to determine one's own learning needs, set goals for education and improvement, participate in appropriate learning activities, and assess the outcomes of education. Effective self-assessment should educate and motivate a physician to engage in further self-improvement educational interventions.[5]
- A critical part of self-assessment includes the identification of one's strengths and areas needed for improvement.
- The literature suggests that we are more accurate in judging our limits at the point of care (reflection-in-action) than assessing past performance (reflection-on-action).
- Discrepancy exists between external assessment and self-assessment, with physicians that perform least well by external assessment demonstrating the greatest difficulties with self-assessment.
- Recognition of the discrepancies between self-assessment and assessment by others and conscious development of self-assessment skills during residency training may aid learners in improving their self-assessment before they enter unsupervised practice where external feedback is more difficult to obtain.

References

1. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher* 2006; 28:117-128.
2. Eva KW, Regehr G. Self-assessment in the health profession: A reformulation and research agenda. *Academic Medicine*. 2005; 80(10 suppl):S46-54.
3. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence. *JAMA* 2006; 296:1094-1102.
4. Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, McKinstry B. The effectiveness of self-assessment, on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide No. 10. *Medical Teacher* 2008; 30:124-145.
5. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine*. 1991; 66(12):762-769.
6. Hojat M, Nasca TJ, Erdmann JB, Frisby AJ, Veloski JJ, Gonnella JS. An operational measure of physician lifelong leaning: Its development, component and preliminary psychometric data. *Medical Teacher* 2003; 25:433-437.
7. Schon D. *The Reflective Practitioner. How professionals think in action*. London: Temple Smith; 1983.
8. Eva JW, Regehr G. Knowing when to look it up: A new conception of self-assessment ability. *Academic Medicine*. 2007; 82(10 suppl):S81-S84.
9. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education* 2008; 42:468-475.
10. Galbraith R, Hawkins R, Holmboe E. Making self-assessment more effective. *Journal of Continuing Education in the Health Professions* 2008; 28(1):20-24.
11. Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: Implications for self-directed learning. *Advances in Health Sciences Education* 2006; 11:235-244.
12. Li ST, Paterniti DA, Co JP, West DC. Successful self-directed lifelong learning in medicine: A conceptual model derived from qualitative analysis of a national survey of pediatric residents. *Academic Medicine*. 2010; 85(7):1229-1236.

Annotated Bibliography

Eva KW, Regehr G. Self-assessment in the health profession: A reformulation and research agenda. *Academic Medicine*. 2005; 80(10 suppl):S46-54.

This paper addresses the tension between self-assessment as a critical part of professional regulation and the lack of evidence for effective self-assessment. The authors define self-assessment broadly as the involvement of learners in judging whether or not learner-identified standards have been met. They highlight the importance of this assessment in identifying an individual's strengths and weaknesses and how reflection on these can enhance one's clinical care of patients. However the complicated process of self-assessment involves a number of interacting cognitive processes and functions as "a monitor, a mentor and a motivator through processes of evaluation, inference, and prediction". They stress that to achieve improvements in one's performance, one must seek out external feedback from reliable and valid sources. Ongoing monitoring and reflection on one's performance is critical in the process of improvement. Overall, this is a useful article that defines the terminology and the controversies in the use of self-assessment. It provides a nice discussion of reflection-on-action versus reflection-in-action for program directors.

Hojat M, Nasca TJ, Erdmann JB, Frisby AJ, Veloski JJ, Gonnella JS. An operational measure of physician lifelong leaning: Its development, component and preliminary psychometric data. *Medical Teacher* 2003; 25:433-437.

This article identifies lifelong learning as a complex process emphasized in the medical profession. Despite this emphasis we lack a universally accepted definition of lifelong learning and do not have a psychometrically sound instrument to measure lifelong learning. This study sought to develop such a tool, to identify its underlying components, and to assess its psychometric properties. Using a review of the literature and the results of two pilot studies, the authors created a 37 item questionnaire. They performed psychometric analysis on the responses by 160 physicians and included 19 items in the Jefferson Scale of Physician Lifelong Learning (Jeff SPLL). Factor analysis was performed and identified five meaningful factors that were reportedly consistent with the definition and major features of lifelong learning. These factors were need recognition (cognitive aspect), research endeavor (capabilities), self-initiation/self-directed learning (behavioral aspect), technical/computer/skills (skills), and personal motivation (predisposition). Validity and reliability of the factors were assessed and the authors concluded that lifelong learning is a multifaceted concept. This article is useful for program directors because of the pertinent discussion of lifelong learning as well as the tool itself. A subsequent publication by Hojat, Veloski, and Gonnella (*Academic Medicine* 2009; 84(8): 1006-1074) examined the psychometric properties and correlates of the revised Jeff SPLL tool and found it to be psychometrically sound and measured physicians' orientation toward lifelong learning among fulltime clinicians and academic clinicians.

Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence. *JAMA* 2006; 296:1094-1102.

This systematic review compares the accuracy of physician self-assessment with external observations of their competence. Searches of multiple databases between 1966 and 2006 using the terms "self-directed learning", "self-assessment", and "self-reflection" yielded over 700 articles, of which 17 met the study inclusion criteria. The authors conclude that physicians have limited ability to self-assess. In addition, those physicians who were in the lowest quartile and were the least skilled by external assessment also had poor self-assessment skills. Through their analysis they defined three discrete types of self-assessment including predictive, summative, and concurrent. They conclude by suggesting that new processes need to be developed to aid practitioners in the self-assessment process. This article is a helpful review addressing methods and accuracy of self-assessment.

Violato C, Lockyer J. Self and peer assessment of pediatricians, psychiatrists and medicine specialists: Implications for self-directed learning. *Advances in Health Sciences Education* 2006; 11:235-244.

The purpose of this study was to examine the discrepancy between self and peer assessments for physicians from several specialties, including pediatricians, internal medicine specialists, and psychiatrists. Data was collected from 304 practitioners in Canada and each individual received assessments from 25 patients, 8 medical colleagues, and 8 non-medical co-workers. Raters used a five point scale for assessment. Data analysis compared the self-assessments with medical colleague assessments on percentile rankings. The overall results provided strong evidence that physicians perform poorly at self-assessment. Specifically physicians in the lowest quartile over-rated themselves compared with peers and conversely those in the highest quartile rated themselves lower than did their colleagues. This is problematic especially for physicians having completed residency training because the opportunity for feedback declines once in practice. The discussion of this paper provides information on the problems with self-assessment accuracy. However the work was done looking at three specific residency training programs and it is unclear if the results can be generalized to all physician specialty groups.

Eva JW, Regehr G. Knowing when to look it up: A new conception of self-assessment ability. *Academic Medicine*. **2007; 82(10 suppl):S81-S84.**

Acknowledging that self-assessment is noted to be poor when compared to external assessment, the authors studied the validity of a new conceptualization of self-assessment in practice. Results revealed that participants showed behavioral indications of being aware of the limits of their ability. This awareness occurred "in the moment" of the edges of their knowledge and competence. It highlights that using reflection-in-practice is a valuable tool. They were able to show that participants knew what they did and did not know. They hypothesize that people have the capability to slow down and research items they need more knowledge about in order to make decisions which should aid in the delivery of safe patient care. Use of these behavioral measures shows that self-assessment when used as an ongoing monitoring process is helpful in daily practice.

Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education* **2008; 42:468-475.**

By engaging in reflective practice, physicians may become more conscious of their current reasoning process. Evidence for this practice is limited and the authors conducted an experiment to study the effects of reflective practice on diagnostic accuracy. Their findings support the concept of reflective practice improving diagnoses made by physicians in situations of uncertainty and uniqueness. This is a critical finding as such reflective practice can lead to a reduction in diagnostic errors. This paper includes an excellent discussion about the behaviors and reasoning processes involved in reflective practice. Reflective practice is encouraged for physicians so that they can critically think about their reasoning and decisions involving patient care.

5. Portfolios in Medical Education

Carol L. Carraccio, MD, MA

"At present our assessment methods stem from the reductionist philosophy that underpins our discipline, and we are, thus, trapped by our need to compare like with like. Until we can make a mental shift that allows us to include a more holistic approach to assessment, one which values the development of individuals over time, we will continue to struggle to measure the unmeasurable, and may end up measuring the irrelevant because it is easier." [1]

Rationale

With the introduction of the six broad and diverse, yet overlapping, domains of the Accreditation Council for Graduate Medical Education (ACGME) competencies, program directors grappled with how to teach and assess these competencies in meaningful ways. Interest in the graduate medical education (GME) community about the possibility of using portfolios for assessment of the ACGME competencies was actually stimulated by experience with portfolios outside of the realm of medical education in the United States. In the early days of the Outcome Project, the ACGME created a "toolbox" to help guide program directors in assessing the competencies; a portfolio was suggested as one assessment tool for difficult to measure competencies like practice-based learning and improvement. Since that time one school of thought about portfolios has evolved from seeing them as a tool to assess an individual competency to seeing portfolios as a system of assessment. For the purposes of this chapter, a GME portfolio is defined as an assessment system that contains evidence of progression towards proficiency in the ACGME competencies and consists of two components: learner selected items (*unconstructed component*) such as best work products and reflections, and evidence from an array of qualitative and quantitative assessments (*constructed component*) chosen by the program director.

Goals

1. Develop familiarity with a working definition of "portfolio" as a system of assessment.
2. Understand the common elements of competency-based medical education and a portfolio system of learning and assessment.
3. Recognize the importance of reflective practice as an added value of portfolios.
4. Learn about and weigh the advantages as well as the limitations of portfolios.

Case Example

It is time to orient a new group of interns. This year you decide to include a session on the new portfolio assessment system that you introduced last year. Over the course of the year you spent a lot of time answering questions from both residents and faculty and doing one-on-one training sessions which makes you think that you need to better prepare incoming interns as well as provide more faculty development.

In order to introduce portfolios to the interns in a way that is meaningful, you decide to assess the interns' understanding of portfolios and you meet with them as a group to engage them in a needs assessment. A number of interns know only of artist portfolios and do not understand their use in the context of GME or what might be included in their portfolio. A few interns who had used a form of electronic portfolios in medical school raise a concern about who will have access to it. Some want to know why you chose to use a portfolio system of assessment to measure the ACGME competencies.

You found this feedback to be helpful and as a result sent an email to faculty mentors telling them that you plan to do an update on the portfolio system for them and ask what topics would be of most interest now that they have had a chance to use the system over the last year. The feedback from faculty, many of whom are mentors for the residents, encompasses some of the same questions that arose from the interns and in addition, they also want to discuss some of the challenges of portfolio assessment, including some of the psychometric challenges.

Points for Consideration

What is a portfolio, what would it contain, and how would it be used in the context of GME?

The word "portfolio" conjures up a variety of images. At one end of the spectrum there is the artist's portfolio, the contents of which are determined solely by the artist. It is typically a compilation of the artist's best work. At the other end of the spectrum, portfolios can be compilations of required information, which is structured both in form and content, and that may have limited input from the learner in its assembly. Between these two extremes, there are many hybrids and variations. However, in each of these cases, the portfolio is seen merely as a collection of work products or assessments or some combination of the two. To move beyond a repository to a system of assessment, work products and assessments are necessary but not sufficient.

A system is a "regularly interacting or interdependent group of items forming a unified whole" (Webster's Ninth New collegiate Dictionary). In the context of a portfolio, the unified whole is a comprehensive perspective on a learner's progress. For our purposes, in this chapter, a *portfolio* system of assessment requires: (1) interaction between the learner and a mentor, each playing an active role, (2) multiple assessment methods including reflection and self-assessment that when aggregated contribute to a comprehensive assessment of the learner, and (3) a longitudinal perspective that demonstrates growth over time.[2]

As mentioned above, the portfolio consists of two components: (1) learner selected items such as best work products (e.g., a PowerPoint presentation of a talk, a research abstract) and reflections (e.g., essay on an ethical dilemma, journal entry on a particular patient encounter), and (2) evidence from an array of qualitative and quantitative assessments of competence in the six ACGME domains.

Portfolios can be either paper-based or electronic. The latter provides for greater practicality, particularly if web-based, because access is readily available at any time. There are a number of institutions and proprietary companies that have created web-based platforms for portfolios. Using one of these systems allows program directors to focus on content rather than technology.

Why would you want to use a portfolio system to assess the ACGME competencies?

Competency-based medical education (CBME) and portfolio assessment have several critical features in common which creates a synergy when a system of portfolio assessment is used to evaluate competence.[3] A fundamental feature of both is the active role that learners must play in driving not only their learning but also assessment of their learning. In fact, assessment becomes the "teachable moment" when portfolios are used.[4]

Another of the underlying principles of CBME is that multiple methods of assessment by multiple assessors are needed to assess the progression towards clinical competence. Web-based portfolios provide a realistic and manageable method of orchestrating multiple methods of assessment as well as multiple assessors while providing a transparency of process that is important to learners and faculty.

Finally the uniqueness of the individual learner is at the heart of both CBME and a system of portfolio assessment. Establishing a personal trajectory for learning and assessment during education and training provides a strong foundation for instilling the habit of continuous professional development over a career.

How do portfolios actively engage learners in reflective practice and what is the benefit to the learner?

Portfolios lend themselves to both quantitative and qualitative assessments.[5] The ability to assess what really matters in judging whether someone is a good doctor forces us to go beyond simple Likert scales of *global assessments* and checklists and embrace qualitative measures. The integration of the quantitative and the qualitative provides a more comprehensive or holistic look at the learners and their capabilities.

Portfolio learning and assessment also prepares the learner for the American Board of Pediatrics (ABP) Maintenance of Certification (MOC) process. Upon passing the initial certification examination, one automatically enters MOC. The ABP provides each diplomate with a personal web-based portfolio for engaging in the four components of MOC. The use of a portfolio in GME helps to set the stage for a more meaningful MOC experience.

As physicians we enjoy the privilege of being a self-regulating profession. But with this privilege comes the responsibility to be accountable to the public and our profession. In essence, this means that we must reflect on our practice for the purpose of continual improvement in care delivery. The process of reflection is not necessarily an intuitive one. The portfolio provides the learner with a stimulus and a platform for self-initiated and guided reflection.[6] Examples of reflective activities may include completing self-assessment of performance, selecting a work product as evidence of achievement of competence in a particular domain, or writing an essay or journal entry addressing an ethical or professional dilemma, to name a few.

An example of an activity involving guided reflection is the semi-annual review of evaluations that takes place between a resident and program director or advisor. In preparation for this meeting, a written template that walks learners through a review of their portfolio and prompts them to answer specific questions about their overall performance, strengths, areas of needed improvement, learning goals and individual learning plans, challenges, and successes serves as a guide to meaningful reflection on practice that can contribute to professional formation of residents. Guidance is also important because we are not accurate in self-assessing (see Chapter 4). Ultimately one's mature sense of professionalism would drive these reflective activities.

Reflection on developmental progress is likewise an essential component of portfolio assessment, which requires the learner to provide evidence of this progression. Development over time is predicated on formative feedback and formative feedback is more likely to change behavior if it is longitudinal in nature.[7] Many portfolios provide a mechanism for this ongoing dialog through "instant feedback". The latter is typically a threaded discussion that can be initiated by either the resident or a faculty member who wishes to give or request feedback. The ease and efficiency of engaging in this exchange promotes coaching and course-correction.

The process of reflection serves several purposes: (1) involving the residents more actively in their own education and assessment through self-assessment, goal setting, and seeking resources and learning activities that will ultimately improve practice, (2) modeling the critical role that PBLI plays in one's professional development by demonstrating the sequence of asking provocative questions that lead one to reflect on- and for-action, and (3) engaging the resident in a process of self-assessment and goal setting that will be instrumental in laying the groundwork for life-long learning, MOC, and practice improvement.

Who has access to the portfolio?

In any discussion of portfolios learners raise the question of who has access. They are rightfully concerned about writing reflections that are deeply personal and having them visible for anyone to read. The technology exists for individual residents to make choices about what they will share and with whom and what they will not share. In fact, it is critical to have this discussion with trainees as they open their portfolios for the first time. They need to understand that as a program director, required evaluations must be shared with you and their advisor since you are responsible for ultimately verifying their clinical competence to sit for their American Board of Pediatrics (ABP) certifying examination and their advisor will help to guide them and be a resource for them throughout their training. In addition, reflections that you require such as those linked to the semi-annual review will need to be shared. However, when residents initiate a reflection on a difficult case to help work through their thoughts and feelings, they may choose whether or not to share and with whom. It is important to remind residents that personal patient identifiers need never be included in these reflections or other portfolio documents.

Occasionally the bigger question of discoverability by those outside of the health profession, such as lawyers engaged in malpractice claims, is raised. There are no guarantees of protection with portfolios just as there are no guarantees of immunity with any other form of documentation. The peer review statutes in each state will be a determining factor in whether a portfolio is discoverable. Maximum protection will be insured when a formal institutional process designates the portfolio as a peer review document. It is also important to have a formal institutional policy that addresses file access, content and retention. The literature on security/ethics of portfolios per se is sparse. One chapter in an on-line reference entitled "The Resident File" suggests that summative evaluations, dates of training and types of training experiences including procedures, along with records of disciplinary actions, materials required by your specific ACGME Residency Review Committee and any other documents that the program feels are important should be permanently retained.[8] In portfolio language this means that they would be archived once the resident is no longer active within the training program. In that same article the authors point out, however, that formative evaluations and notes about successfully mediated problems may not be necessary to keep unless the program director feels that a similar issue may arise in the future or a lingering question about performance during training will require the given substantiation. There is also the issue of ownership of one's own data. If we believe learners own their data then archiving it in a way that they have access is the right thing to do. This is an area where expert guidance is needed. The ethics and security of portfolios are fertile ground for expert consultation and debate as we are still too early in their use to realize all of the ethical and legal ramifications.

What are other challenges in using portfolios?

The most frequently verbalized criticism about portfolios is that they are too time intensive to be practical within the demanding and fast-paced world of health care and within the context of duty hour limitations. It is not the portfolio that creates the burden of work, but rather the need to assess additional competencies beyond patient care and medical knowledge. A portfolio assessment system facilitates more meaningful judgments about competence in these additional domains because it captures quantitative and qualitative data.[8] A web-based portfolio can relieve the burden of sending and collecting all of the paper evaluations that are needed to address all of the competencies. As with all web-based systems, there is an initial investment of time in getting a portfolio system up and running, and ongoing administrative support is essential. A critical ingredient is a champion who will help to carry a program through the difficult start-up times and encourage a cultural shift from a toolbox philosophy to an assessment system philosophy in order to truly realize the power of portfolios in medical education. This is especially true for web-based systems where the technology sometimes facilitates and sometimes challenges the ultimate goal of efficacy and efficiency.

Any new assessment method requires faculty development and resident orientation. Whether or not you use a portfolio, faculty development regarding assessment of the competencies is crucial since education in assessment is currently not a standard part of the development of residents, fellows and faculty (see Chapter 3).

What are the psychometric challenges?

Although the challenges of scoring may intimidate the potential user of portfolio assessment there are threats to validity with every assessment. The gains in measuring what is "authentic" or relevant to real world practice may outweigh the need to struggle with reliability in scoring the portfolio as a whole. In fact, this characteristic of *authenticity* contributes to the predictive validity of the portfolio.[4] The table below, adapted from suggestions of Tekian and Yudkowsky,[9] lists some of the psychometric challenges.

Psychometric Properties	Challenges	Solutions	Opportunities
Validity	Construct underrepresentation: inadequate sampling	Design or blueprint that ensures a systematic approach for including content from each domain being assessed	An electronic portfolio makes this sampling much easier and more practical considering the 6 broad and diverse ACGME competencies
	Construct irrelevant variance: scores based on more elements than the one being assessed (e.g.. reflections clouded by poor writing skills or reluctance to be honest in addressing weaknesses)	Designate reflections for formative as opposed to summative assessment	Development of competence depends on continual formative feedback and learning; activities requiring reflection present the perfect opportunity to engage in meaningful formative feedback
Reliability	Rater disagreement	Standardize content and use multiple raters when possible

Rater training and calibration through faculty development

Score the components independently and focus on the individual elements

Employ methods of triangulation (combining different assessments of the same ability) and dependability (documenting the assessment process and insuring that process is being followed) | Embracing qualitative assessments adds to the comprehensive nature of learner assessment and likely the value since some very meaningful elements of the competencies do not lend themselves to quantitative measures |

When one uses a portfolio system of assessment there is an added challenge of dealing with contents that are unconstructed and chosen by the learner as well as those that are constructed and added by the program director. For the former, there may or may not be measures available to assess them. For example, if the resident is given the assignment of writing a reflection on an ethical dilemma there may be some key principles that you would expect to emerge in the essay and could give formative feedback to the resident if they are missing. But this work would not be "graded." Its true value lies in the process of reflecting on the encounter and organizing thoughts to put to paper as well as in the formative feedback given by the faculty member who reads it. For the constructed tools, the first step is to decide whether each component of the portfolio will be scored separately and then averaged (*compensatory scoring*), or whether each

component will be scored against a minimum threshold for passing (*conjunctive scoring*). Another alternative is to score the portfolio as a whole.

One reasonable approach to addressing the psychometric challenges is to balance the constructed and unconstructed components of the portfolio. The former will lend to the reliability of the portfolio for summative assessment purposes and the latter often adds to the meaning of what we are measuring. Some learning activities may not be measurable, such as the process of reflection and self-assessment. Measuring the accuracy of this process is not as important as the learning that is gained by engaging in it. Using multiple tools to assess difficult-to-measure skills or competencies and providing faculty development to address the process and use of tools with frequent checking to insure proper use will enhance the *credibility* and *dependability* of these tools.[10] The balance between structured and unstructured components will enhance the ability to sample all the competencies, increasing the evidence that supports the validity of the portfolio.

Lessons Learned

- The value of the portfolio lies in its focus on the whole rather than just the sum of parts or sum of competencies.
- The quality of the feedback that can be derived from portfolio review is dependent on the quality of the learning activities and tools that are built into the portfolio.
- Portfolios can provide a system of learning and assessment; engaging in the process of assessment is an important learning activity.
- Faculty mentoring of residents is critical in maximizing the value of portfolio learning and assessment, particularly as it pertains to reflection and practice improvement.
- The greatest benefit of portfolios is that they are learner-centered, that is, they engage the learner in playing an active role in education and assessment, an essential ingredient for professional formation as it relates to competency-based education/training and maintenance of certification.

References

1. Snadden D. Portfolios: attempting to measure the unmeasurable? *Medical Education* 1999; 33:478-479.
2. Challis M. AMEE medical education guide no. 11 (revised): Portfolio-based learning and assessment in medical education. *Medical Teacher* 1999*;* 4: 437-440.
3. Holmboe ES, Davis M, Carraccio C. Portfolios. In: Holmboe ES, Hawkins RE. *A Practical Guide to the Evaluation of Clinical Competence*. Philadelphia: Mosby; 2008.
4. Friedman Ben David M, Davis M, Harden R, Howie P, Ker J, Pippard M. AMEE guide No. 24. Portfolios as a method of student assessment. *Medical Teacher* 2001; 23:535-551.
5. Carraccio C, Englander R. Evaluating competence using a portfolio: a literature review and web-based application to the ACGME competencies. *Teaching and Learning in Medicine* 2004; 16: 381-387.
6. Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007; 82:493-502.
7. Archer J. State of the science in health professional education: effective feedback. *Medical Education* 2010; 44:101-108.
8. Association of Program Directors in Internal Medicine. The Resident File. In*: A Textbook for Internal Medicine Education Programs* (10th edition). Uniondale, NY: Association of Program Directors in Internal Medicine; 2011.
9. Tekian A, Yudkowsky R. Assessment Portfolios. In: Downing S, Yudkowsky R. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009.

10. Tohel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Medical Teacher* 2009; 31:299-318.

Annotated Bibliography

Challis M. AMEE medical education guide no. 11 (revised): Portfolio-based learning and assessment in medical education. *Medical Teacher* 1999; 4: 437-440.

This comprehensive and frequently cited article provides answers to the following questions: What is a portfolio? Where did portfolios come from? What is the educational rationale for using portfolios? What is the link between portfolios and professional development? What does a portfolio look like? How are portfolios assessed? What are the major issues in portfolio-based learning and assessment? Examples of portfolios used for medical education, graduate medical education, general practice vocational training, specialist registrars, general practice trainers, and continuing professional development are described. A very useful table (guide to the stages of portfolio development and review/assessment) lists 8 tasks, how to do each and who needs to be involved. This is particularly helpful in framing your approach and getting started with a portfolio system of assessment.

Carraccio C, Englander R. Evaluating competence using a portfolio: a literature review and web-based application to the ACGME competencies. *Teaching and Learning in Medicine* 2004; 16: 381-387.

A MEDLINE search (1996-2002) for English language articles on medical portfolios/assessment was conducted, resulting in 35 articles meeting inclusion criteria. Major conclusions were: learners play a pivotal role in driving competency-based education and portfolio-based learning and assessment; formative feedback is critical to the achievement of competence; reflection is critical to professional development and portfolios must balance reflective components with structured evaluation components. A web-based portfolio, developed and implemented, based on the reviewed literature is described. The portfolio is designed to evaluate performance in the six ACGME competency domains and consists of both structured and unstructured or reflective components. Each component is described in detail.

Dannefer EF, Henson LC. The portfolio approach to competency-based assessment at the Cleveland Clinic Lerner College of Medicine. *Academic Medicine* 2007;82:493-502.

The authors describe their experience with designing and implementing a portfolio system of assessment for UME that is framed around the achievement of 9 competencies: research, basic & clinical science, medical knowledge, communication, clinical skills, clinical reasoning, professionalism, personal development, health care systems, and reflective practice. The key features of their assessment system are: (1) the learners drive it, (2) mentors are involved and engaged at every step of the way, (3) the evaluation templates use narratives that identify areas of needed improvement and reinforce strengths, (4) the templates are supplemented by multiple other assessment methods such as an OSCE, knowledge tests, observed H & Ps, etc., (5) work products of the students are used as authentic evidence of competence, (6) ongoing and intense faculty development, and (7) learners are responsible for preparing and presenting their portfolios for formative and summative assessments. The process described for summative portfolio review merits attention. There is a 2-step standard setting process. Each committee member reviews the same sample of 8 portfolios and then meets as a group to come to consensus regarding which standards are essential for demonstrating achievement of each of the competencies. The intent here is to improve inter-rater reliability. For step 2 the committee discusses each of the 8 portfolios and, standard by standard, votes on whether the student has met that standard i.e., achieved competence in that domain. The outcomes for the sample of students are then reviewed to determine "whether cut points are acceptable". At

this point two reviewers are assigned the remaining portfolios for independent assessment. The two reviewers then attempt to reach consensus. Their deliberations are presented to the whole committee. If consensus is not reached by the reviewers or they recommend dismissal, the portfolio is read and voted on by every committee member.

Friedman Ben David M, Davis M, Harden R, Howie P, Ker J, Pippard M. AMEE guide No. 24. Portfolios as a method of student assessment. *Medical Teacher* **2001; 23:535-551.**

This extensive article provides background for the use of portfolios for assessment, reviews the range of assessment purposes for which portfolios have been used, identifies possible portfolio contents and discusses advantages of portfolios particularly for assessing professionalism. Psychometric issues in the context of portfolio assessment are discussed (formative/summative; qualitative/quantitative; personalized/standardized; authentic; rater consistency; decision consistency; sampling; consensus approach; forms of validity; external versus internal examiners). Issues to consider in portfolio implementation are also discussed (defining the purpose, determining the competencies to be assessed; selection of portfolio material; developing a marking system; selection and training of examiners; planning the examination process; student orientation; developing guidelines for decisions; establishing reliability and validity evidence and designing the evaluation procedures).

Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Medical Teacher* **2009; 31:299-318.**

This systematic review on portfolios included 56 articles (27 of which were in medicine). Thirty-three of the 56 study designs were characterized as controlled observational studies limiting the strength of evidence to support any conclusions. Another major problem in the literature is the lack of clear description of the definition of a portfolio, which can be quite variable. Having said that, there are a few take home messages from the authors based on their review: (1) support of a mentor trained in portfolio assessment is a critical ingredient of portfolio success, (2) there is some evidence to suggest that portfolio users are more actively involved in their learning, (3) there are mixed messages about portfolios supporting reflection but then stifling reflection if assessment of the reflection is attempted, (4) the problems with reliability which can be addressed with multiple trained assessors and triangulating portfolio data with other methods of assessment, and (5) the discrepancies about validity that cannot be reconciled are likely due to the very limited numbers of studies that look at outcomes of portfolio assessment. This article does a good job of laying out the limitations to studying the impact of portfolios in medical education.

Holmboe ES, Davis M, Carraccio C. Portfolios. In: Holmboe ES, Hawkins RE. *A Practical Guide to the Evaluation of Clinical Competence***. Philadelphia: Mosby; 2008.**

This is a very readable review of the definitions of portfolios, their purpose and strengths, limitations, and psychometric challenges. In regard to the latter, the authors propose strategies for the creation, content, and implementation of portfolios that will mitigate the challenges. Types of assessment tools to address the ACGME competencies are discussed. In addition, the issue of formative and summative uses of portfolios is debated. The authors make the case for the longitudinal and developmental emphasis that portfolios provide in learner assessment. There is an annotated bibliography of the articles that were important to informing the writing of this chapter as well as an extensive reference list.

Tekian A, Yudkowsky R. Assessment Portfolios. In: Downing S, Yudkowsky R. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009. p. 287-302.

The authors address the various types of portfolios and the content that would be included in each. There is an important discussion on assessing and evaluating the portfolio itself. Threats to validity and issues of reliability are also identified and explained. The table that includes this information is particularly helpful. The chapter concludes with case examples of portfolios that are used for different purposes, addressing what might be included and how they might be scored.

6. Program Evaluation

Stephen Ludwig, MD

"The only thing experience teaches is that experience teaches us nothing." - André Maurois

Rationale

This primer focuses primarily on the assessment of individual trainees. The term evaluation refers to the process of obtaining information about a course or program of teaching for purposes of subsequent judgment and decision making.[1] The sum total of individual resident assessment will also provide indicators of performance of the overall program.

This chapter, unlike the others in the primer, focuses on the program as a whole and how to evaluate it. Training programs have many different parts. You want to make sure that all the parts are fitting together to make the program what it needs to be. At times this seems an impossible goal. However, it is also exciting to know that your work is never done and for many program directors this is the joy of the process. No program is perfect. Every program can be improved.

As a program director you are held accountable for the success of your program. The primary motivation for success is your own sense of pride and professionalism that you are doing a good job preparing physicians to provide effective and knowledgeable health care for children. In addition, there are the individual assessments of those who are training in your program. Do they feel your program is providing a supportive educational environment? Do they feel that the program is helping them to meet their goals? Are the faculty satisfied with the trainees your program is graduating? Are the trainees prepared for practice, for fellowship, or for life as a pediatrician? Are you providing an atmosphere for training that is advancing the mission of the institution? In addition, there is public accountability through external accrediting bodies such as the Accreditation Council for Graduate Medical Education (ACGME), which will assess the success of your program in meeting minimal standards for accreditation, and the American Board of Pediatrics (ABP), which will assess and certify the competence of your trainees.

With so many possible constituencies and stakeholders in both the process and outcome of your program, you will want to make sure that your program is meeting its goals and objectives. If it is not, you will want to make some course corrections to place your program in a better position to do so.

Goals

1. Understand the purpose and the process of *program evaluation*.
2. Describe methodologies and strategies for program evaluation.
3. Recognize external requirements and internal requirements for program evaluation.

Case Examples

Case 1

One of your graduated residents contacts you to tell you that she was just notified that she failed the ABP Certifying Examination. You are surprised as your memory of this resident is that she was an excellent clinician and a bright and energetic young woman. You try to console her and help her work through some strategies to pass the next time. You make plans to follow up with her to help her in this goal. After

your conversation is concluded you begin to think, "I wonder if there are others like her? Did our program fail her in some way? How do we know that this program is providing the right training?" We cannot assume that this failure is only the failure of an individual; perhaps it is a program weakness.

Case 2

A graduate of the program reports that she has entered a primary care practice and she does not feel well prepared. She reports feeling comfortable with managing acute illness but very uncomfortable with behavioral/developmental issues. One of the senior practice partners seems to do this work so well. He has suggested she take a CME course. Is this something that requires more time and experience in the context of on-the-job training? Should the program be responsible for meeting the needs of every trainee and every career path?

Case 3

You bump into the fellowship program director from the nephrology division in the cafeteria. He complains that none of "your" residents are going into nephrology. His impression is that it must be the fault of the program. He waxes eloquent about the small number of subspecialists nationally and urges you to do something about it. Is the lack of nephrologists your problem to fix? What does his reference to "your" residents mean? Is there a relationship between residency experience and career choices?

Points for Consideration

Does your program have defined goals and objectives? What is your program trying to accomplish?

Just as any single rotation or smaller segment of an educational program needs to define goals and objectives in order to chart its course and to establish evaluation standards, a program as a whole also needs goals and objectives. Moreover, ACGME requires the program leadership, Department Chair, and faculty to set out goals and objectives for the program.

Clearly, the overarching goal of a residency program is to train pediatricians who will have successful careers providing safe and effective care. However, most programs have other goals and objectives as well. For example, "The goal of our program is to train academicians in subspecialty careers" or "Our goal is to train good primary care pediatricians who will stay in our region". Articulating these goals is very worthwhile, as success must be measured against some standards, and program goals and objectives establish these standards. Failure to articulate goals can also leave some *stakeholders* feeling the program is on target while others feel it is missing the mark.

Who defines success of a program? Who are the stakeholders?

Trainees: Perhaps the most important group to define success of your program will be the trainees themselves. Current trainees can provide prospective data on the program; past trainees can provide retrospective evaluations. Prospective evaluations by current trainees may be influenced by factors such as fatigue, stressful work, and their self-perceived low status within the hospital. Nonetheless, these evaluations have relevance for what is needed for immediate change and may be an important factor in recruitment. Retrospective evaluations at 3-5 years post-training are also important as the aforementioned stressors may have dissipated and "real-life" career factors can be assessed (as in Cases 1 and 2). Unfortunately, the long term view removes the possibility of making program corrections for those who have already graduated.

Program Director: The program director also defines success—not only of the program but of her role. Some program directors find that after 4-5 years the job is no longer satisfying. Other program directors seem to embrace the role for 10, 15 or more years. It is important to make an annual assessment of your career satisfaction as the PD.

Faculty: The faculty and Department Chair also are likely to measure success of a program. Their assessment is often linked to their sense of involvement and to the goals and objectives of the program (as in Case 3). Sometimes a faculty member that has evaluated the program negatively expresses his view through disengagement with the program. Sometimes their critiques are more directed to specific issues. The ACGME requirement that programs engage faculty in at least annual evaluations and action plans are aimed at eliciting this kind of faculty feedback.

Sponsoring Institution: The sponsoring institution may have its own measure of success or failure of a program. There needs to be an assessment by the institution as to how well the program is meeting the hospital's goals. Institutions often express their assessment in terms of allocation of resources to the program. If resources are available but not available to the program it may be based on a negative institutional assessment. The institution will also evaluate the program through the office of the Designated Institutional Official (DIO) and the Graduate Medical Education Committee.

ACGME: The ACGME is the organization charged with assuring the public that programs are meeting their responsibilities to the trainees and to society. Every program director is aware of the ACGME evaluation methods, including the Confidential Resident Survey, the site visit and Site Visitor's Report, and the review of documentation by the Pediatric Review Committee (RC). The sum total of these data results in an accreditation status and a cycle length that lets a program know where it stands relative to national requirements.

Other external stakeholders: There may be other external stakeholders that evaluate your program. Other licensing organizations, the Joint Commission, State Health Boards, specialty boards, and others may have made evaluations relevant to your program. As a program director it is up to you to determine which of these sources of evaluations are reliable and valid. You will also need to determine how transparent the results of your own evaluations should be. Do you report your findings internally, externally, via public announcement, or upon request?

When is the best time to evaluate the program?

Programs should be evaluated from several different vantage points and at several points in time. Because there are many stakeholders in the process, each with its own interpretation of success, it may be impossible to satisfy everyone. Evaluation is a moving target. It is never perfect and there is no best time. So, ongoing program evaluation may be the best course of action.

What approach do I take in evaluating my program?

Kirkpatrick and Kirkpatrick, in their landmark publication on "Evaluating Training Programs", discuss four levels of evaluation: (1) evaluating reactions, (2) evaluating learning, (3) evaluating behavior, and (4) evaluating results. [3] Goldie summarizes the history of program evaluation and provides a framework for potential evaluators. His paper covers the role of the evaluator, ethics of evaluation, and evaluation design and implementation.[1]

One way of characterizing evaluation approaches is to consider the role of *process evaluation* and *outcome evaluation* in your program. Process evaluation is directed by the question, "What is the program doing?" Outcome evaluation answers the question, "What has the program done?" Both are valuable.

In process evaluation, you consider the program's activities. You might be interested in determining what kind of resources your activities require, whether the activities cover all of your objectives, and how many learners participate in your activities. Process evaluation will alert you to current issues and problems in the operation of the program.

Outcome evaluation will tell you whether you are truly accomplishing your goals and objectives. Everything might be fine on your process evaluation but if trainees are not passing their certifying examination or finding they are not prepared for their first job, then the program has a problem.

What are the tools to be used for evaluation?

No single evaluation tool does it all.[2] Thus, you must use several different tools. Program evaluation will be a patchwork of different kinds of measurements as described below. It will be a misshapen blanket that may not cover each and every part of your program. But by using many pieces of patchwork, most of your program will be covered. The tools listed below represent patchwork. Different combinations will be used to cover different parts of the program.

What methods or strategies will give me the information that I need?

Surveys: Surveys of residents are a common method of program evaluation.[4-6] They are conducted at many points during a resident's career and they are generally used to assess process. On the surface, surveys appear to be easy to construct but there are several questions to be answered when planning to survey:

- Are you really asking the question you wish to ask?
- How do you know if those who complete the survey are representative of the entire group?
- What form of response will be most helpful? A checklist? A Likert scale? An open-ended response?
- Are all responses considered equal, or are the responses of some residents more important than others?
- How can you increase the likelihood that trainees will give honest and accurate answers?

To be useful, surveys need careful thought and planning. It is often helpful to assemble a focus group or conduct cognitive interviews to review a survey before you administer it. Web-based survey tools (e.g. Survey Monkey®) are readily available, easy to use, and can enhance anonymity.

The ACGME Survey will play an increasingly important role in the ACGME evaluation of a program. Program Directors should be aware of the results of this survey and proactively prepare residents for the survey, correcting any deficiencies that the survey reveals.

Scores: Another evaluation technique is to look at scores. These are often powerful outcome measures because scores are easily quantified and summarized, and facilitate comparisons over time and among groups. Again, some caution needs to be taken.

- What is an appropriate reference group for comparisons? For example, do you compare your residents with a contemporaneous national reference group or do you compare your program's current performance with its past performance?
- Do the scores tell you about individual achievement? For example, the mean In-Training Exam (ITE) score for your program may be high because you have several outstanding residents but you may have a small group who are in need of help.

- Are the scores a reflection of true medical knowledge or test-taking ability? Do they measure problem solving or other important skills needed to be a good physician? In short, do you have evidence of their validity for your intended use (see Chapter 1 for more detail)?

Scores on ITEs or certifying exam pass rates are helpful but have limitations. Examinations take place at one point in time and are focused on the medical knowledge domain of competence. You need other kinds of assessment to get a complete picture of how your program is addressing the other five ACGME competencies.

Products/Portfolios: Another method of evaluation is to look at products or accomplishments of the residents in your program in the aggregate. Some of this information is requested on the ACGME Program Information Form (PIF) and may be collected in the portfolios of individual residents (see Chapter 5). As Program Director, you may wish to create a portfolio of portfolios where you assemble the work products (e.g. papers, lecture handouts, project descriptions) of the entire program and use aggregate data to evaluate the program.

Program 360° or Multi-Source Evaluation: Just as individual residents may undergo a multi-source feedback assessment, the program may benefit from a similar multi-source evaluation.[7] With this method the various program stakeholders have an opportunity to express their evaluations at one time.

Outcome of Recruitment: Another source of evaluation data is your recruitment success. Many factors contribute to recruitment, but the most important are the reputation of the program and what current trainees tell potential candidates about the program. Looking at who matched, and surveying those who chose to go elsewhere, is an important evaluation procedure that can yield useful information. The National Resident Match Program (NRMP) will also provide data about how your match cohort of residents compares to that of other programs.

Systematic Consultations: Another evaluation method is to assemble a group of consultants outside your program to perform a review. This is the approach used in the ACGME mid-cycle internal review, and is also available through the APPD Consultative Program. You could also invite an outside consultant or consultants to visit and perform a review. Fresh eyes and ears can sometimes uncover important issues or factors to which the program director may be blind. External evaluation can be costly, but may be well worth it if it alerts the program director to looming problems and leads to improvement of the program.

Benchmarking: Comparing your program with other programs within and outside your institution can also provide important data for evaluation.[8-10] One example of benchmark data is the National Resident Matching Program results report, which compares the match rate for every participating residency program. Although this strategy has been relatively uncommon to date, it is likely to become more commonplace in the context of medical education. The APPD project Longitudinal Education and Research Network (LEARN) will be helpful in this type of evaluation; the purpose of LEARN is to create research networks that will study educational interventions at residencies nationwide.

What are common barriers to change?

In every improvement process, there are barriers to change. It is important to identify these barriers in order to be able to overcome them.

The most common barrier is inertia. A program director may cling to the notion that "we have a good program" that doesn't need evaluation or change. This can keep program directors from even initiating the program evaluation process. Even an excellent program can become a better program. Finding areas

of weakness will only lead to further improvement. The process of program improvement also serves as a good role model for trainees who will be required to participate in improvement efforts.

A second barrier is the fear that others will assume that a program with deficiencies must be led by a poor program director. Although the program directors are important, there are many factors that contribute to success. Don't feel threatened. In fact, it is the poor program director who does not seek to evaluate and improve her program on a regular basis. Each program must decide what to do with the evaluation data it collects. There will be decisions about the validity of the data and the level of transparency you wish to have.

Program directors may perceive an institutional resistance to change. Sometimes institutions move slowly, but most do move. Gathering program evaluation data is a powerful way to stimulate change especially when you can benchmark your data to that of other comparable programs.

Finally, there are the barriers of limited time and resources for performing a program evaluation. This is where you can seek help from faculty colleagues, graduate students wanting to do a project, or administrative support from your hospital.

Lessons Learned

- You will need a variety of evaluation techniques, including both process and outcome evaluation by multiple stakeholders.
- The many stakeholders in a program may have different agendas and the Program Director must be aware of these.
- The most difficult part of program evaluation is putting it all together. Remember that you will never have a perfect program that satisfies everyone in every circumstance.
- Select and foster an evaluation group that will be honest, critical and constructive. The ACGME suggests using faculty and residents to do an annual review and set out a defined action plan for the coming year. The action plan should contain specifics about the steps to be taken, for example, who will be responsible for implementing the steps? How will you measure or know whether you accomplished the proposed changes? And were the changes appropriate? Be as specific as possible in defining the action plan. Encourage the group to think creatively and not to be restricted by what exists but to engage in expansive thinking about what will be best for the program.

References

1. Goldie J. AMEE education guide no. 29: Evaluating educational programmes. *Medical Teacher* 2006; 28:210-224.
2. DeSilets LD. Connecting the dots of evaluation. *Journal of Continuing Education in Nursing* 2009; 40:532-533.
3. Kirkpatrick DL, Kirkpatrick JD. *Evaluating training programs: The four levels* (3rd ed.). San Francisco: Berrett-Koehler; 2006.
4. Parrino TA, Kern DC. The alumni survey as an instrument for program evaluation in internal medicine. *Journal of General Internal Medicine* 1994; 9:92-95.
5. Seelig CB. Quantitating qualitative issues in residency training: development and testing of a scaled program evaluation questionnaire. *Journal of General Internal Medicine* 1993; 8:610-613.
6. Seelig CB, DuPre CT, Adelman HM. Development and validation of a scaled questionnaire for evaluation of residency programs. *Southern Medical Journal* 1995; 88:745-750.

7. Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. *American Journal of Physical Medicine & Rehabilitation* 2003; 82:394-402.
8. Phitayakorn R, Levitan N, Shuck JM. Program report cards: evaluation across multiple residency programs at one institution. *Academic Medicine 2007*; 82:608-615.
9. Bellini L, Shea JA, Asch DA. A new instrument for residency program evaluation. *Journal of General Internal Medicine* 1997; 12:707-710.
10. Henker R, Hinshaw AS. A program evaluation instrument. *Journal for Nurses in Staff Development* 1990; 6:12-16.

Annotated Bibliography

Bellini L, Shea JA, Asch DA. A new instrument for residency program evaluation. *Journal of General Internal Medicine* 1997; 12:707-710.

This is a report describing the development of a comprehensive program evaluation instrument. The instrument is a 60-item 5-point rating scale. The instrument was given to 104 residents in one program and found to be psychometrically sound, comprehensive, and exportable. There were three sub-classes of response: workload, education, and environment and lifestyle issues. The authors indicate that this instrument may be useful in program evaluation. Based on the use of the instruments the authors were able to make some specific modifications to their program.

Goldie J. AMEE education guide no. 29: Evaluating educational programmes. *Medical Teacher* 2006; 28:210-224.

This guide reviews the history of program evaluation. It provides a framework for potential evaluators considering undertaking program evaluation. The paper discusses the role of the evaluator, the ethics of evaluation, choosing the questions to be asked, evaluation design including the dimension of evaluation and the range of evaluation approaches available, and interpreting and disseminating the findings. Overall, this is a comprehensive look at the program evaluation process.

Parrino TA, Kern DC. The alumni survey as an instrument for program evaluation in internal medicine. *Journal of General Internal Medicine* 1994; 9:92-95.

The authors point out that one principle of continuous quality improvement (CQI) is that the quality of the production process must be measured by the performance of its products. With that principle in mind the study was a survey of a previously published report of 6 alumni groups of Internal Medicine Programs. The authors sought to find whether basic skills and practice-related issues were underemphasized. Their survey included six steps: (1) problem definitions and formulation of objectives; (2) identification of elements for observation; (3) assembling of available studies; (4) assessment of the quality and applicability of each study; (5) comparison of quantitative results; and (6) development of an overall assessment. The authors concluded that such surveys are useful for periodic re-evaluation of a program's goals and outcomes and they fulfill the precepts of CQI.

Phitayakorn R, Levitan N, Shuck JM. Program report cards: evaluation across multiple residency programs at one institution. *Academic Medicine 2007*; 82:608-615.

This paper reports on an instrument potentially valuable to Designated Institutional Officials (DIOs) who may wish to evaluate multiple programs at one institution. It creates a report card that looks at (1) quality of candidates recruited, (2) the educational program, (3) graduate success, and (4) overall house officer satisfaction. The authors conclude that although this technique is useful to DIOs and Program Directors it

is difficult to provide concrete construct validity. The authors recognize that this is a surrogate for the RRCs' perception of quality. Its value may be in the alignment of programs within an institution and as a tool for the DIO.

Seelig CB. Quantitating qualitative issues in residency training: development and testing of a scaled program evaluation questionnaire. *Journal of General Internal Medicine* **1993; 8:610-613.**

This publication describes an effort to develop and test a scaled program evaluation questionnaire focusing on resident satisfaction with workload, learning environment, and stress. The author surveyed 92 internal medicine residents from 5 programs. Residents came from all three years of training. Phase 1 of the study included developing the questionnaire and Phase 2 was the application of the questionnaire over a three year period at a single institution. The authors conclude that this is an exportable instrument.

Part II: Assessment of the ACGME Core Competencies

The "Good Doctor": The whole is greater than the sum of the parts

Carol L. Carraccio, MD, MA

"On one hand, competencies are usually formulated as broad general attributes of a good doctor. On the other hand as soon as we attempt to assess competencies they tend to get reduced to detailed skills or activities."[1]

Rationale

There is great debate in the medical literature about how one evaluates physician competence.[2,3] Prior to 2001, faculty completed global assessments of residents, based on direct or indirect observation over time. These global assessments, along with the American Board of Pediatrics (ABP) In-Training Examination, were used to measure resident performance. Since 2001, the Accreditation Council for Graduate Medical Education (ACGME) has emphasized development of six core competencies during residency and fellowship training.[4] The parsing of physician competence into six component parts prompted educators to break down complex tasks involving knowledge, skill, and attitude and to assess the individual competencies using checklists and other tools.

The assessment of individual competencies may be necessary for assessing resident performance, but it is not sufficient. For example, a resident who performs well when observed doing a history and physical using an assessment tool comprised of an itemized checklist of questions asked and organ systems examined may still not be able to provide optimal care to patients. He may not be able to synthesize the information in a meaningful way to develop a differential diagnosis and management plan, or be capable of explaining the management plan to the patient in a way that can be understood and followed. Thus, the issue is not which type of assessment provides the best method for understanding developing competence, but rather how to blend and balance evidence from different types of assessments, each of which contributes to a comprehensive assessment that relates in a meaningful way to what the physician will be called upon to do in real world practice.[5]

Goals

1. Learn how to provide a more comprehensive and meaningful assessment of learners by balancing the assessment of discrete skills with the holistic assessment of the integrated knowledge, skills, and attitudes skills that one uses in real world practice.
2. Envision the progression to competence as a developmental model marked by milestones achieved along the educational continuum from premedical studies to continuing medical education.
3. Think about the competencies in the context of what a pediatrician does in everyday practice (e.g., caring for a normal newborn) to make them more meaningful to trainees and to faculty.

Case Example

A faculty member sees you at a meeting and pulls you aside to tell you that he is worried about one of the interns "who just doesn't seem to get it". It's now April, so this is particularly worrisome, because you were counting on this intern transitioning to a supervisory role as PGY-2 in a little over 2 months. You go back to the office and pull the resident's file. Until now, although there are no assessments to suggest performance has been excellent, all evaluations have shown a satisfactory performance, with the exception of the mid-year continuity clinic assessment. The latter suggests that the intern is progressing in skill acquisition but at a much slower rate than would be expected. His performance on his in-training exam (ITE) and his structured observations of histories and physical examinations were again noted to be satisfactory. You are upset that this concern is emerging so late in the year and you wonder how to interpret the concerning report; you wonder if there isn't a better way to be informed about a resident's overall performance status or readiness to supervise.

Points for Consideration

_If_ there is a real concern in performance, _why_ did it emerge now?

Perhaps the resident is just now encountering a rotation that is particularly challenging. Perhaps other personal needs are impacting the resident's ability to function at work. Could the problem lie with the tools, how the faculty who use the tools, or some combination of the two? As a community we are still searching for reliable and valid tools for assessing competence. The tools currently in use at the program depicted in the case example may not be reliable or valid, or raters (evaluators) may not be trained to provide standardized assessments.

The tools themselves may have inherent problems. If they do not adequately sample the knowledge, skills and attitudes that are important in the given learning environment, their validity is limited (see Chapter 1).[6] They may be subject to systematic rater bias.[6] For example, one faculty member may be severe in his judgments while another is lenient. In addition, a faculty member may extrapolate from one trait of the learner to other traits. An example of this _halo effect_ occurs when faculty judge a resident's skills in performing a history and physical based solely on his ability to present the case. An articulate resident with organized thinking may be judged as being competent to perform a comprehensive and accurate physical examination, even when he performs maneuvers incorrectly.

There are several ways to reduce rater bias. Ratings by multiple raters, trained at assessing resident performance, will improve the reliability of the assessments. Standardizing the interpretation of the items and the scoring that is used in assessment is critical. For example, a tool that uses a scoring rubric based on "below standards, meet standards, and above standards" is open to wide interpretation unless there is clear documentation of what those descriptors really mean. There are several ways to address this problem. One is to create narrative descriptions of behaviors referred to as _anchors_, and attach them to each score. Videotaping of behaviors to demonstrate what is meant by each of the anchors and using these to train faculty raters will help to standardize scoring by "calibrating" the raters and thus improve both the intra-rater and inter-rater agreement or reliability.[7]

Why is the assessment from the continuity clinic preceptor different from the others?

One of the reasons for this assessment being different from the others is alluded to above: the clinic preceptor may be a more severe rater than the other faculty who assessed the resident. Another consideration that is particularly relevant in the context of our current system of training and patient care is possible erosion of longitudinal relationships between faculty and residents. Faculty often rotate through clinical

services on a weekly basis, pulled in many directions at once, and with little opportunity to observe resident skill progression. In the scenario presented here, the continuity clinic preceptor may be the only faculty member who has observed this resident over time and therefore is the only one that can identify the delayed skill progression.

Finally, there is a construct known as *case specificity*: physicians may not transfer skills learned on one case to other cases.[8] Deficiencies in knowledge as well as differences in the learning environment in which knowledge is applied and the skills are practiced may be responsible. An example of the latter would be a resident who typically encourages flu vaccine for patients with asthma in his continuity clinic but often forgets to ask whether patients being discharged from the hospital during flu season have been immunized. This practice is a habit in the outpatient clinic but may not generalize to the inpatient setting. These concepts are important considerations for faculty who assess resident performance and point to the critical importance of faculty development (see Chapter 3), particularly for clinician-educators.

How can you appreciate where a trainee is in the developmental sequence, so that you can facilitate their progression; and how can you identify learners who are struggling and intervene early so that you can provide optimal chances for remediation?

Just as the Denver Developmental Screening Test[9] is helpful in identifying children at risk of delay, there are some developmental models of *professional formation* that may be helpful in the early identification of learners whose skill progression is not occurring at the expected rate. Like the Denver, such models provide a description of the developmental process and provide the resident with a learning roadmap. For the few residents who struggle, use of such models allows early identification of difficulties and opens the possibility of early intervention and timely remediation, thus addressing a learner's self-esteem and self-efficacy.

The Dreyfus and Dreyfus model describes a skill progression from novice, through advanced beginner, competent, proficient, expert and master.[10,11] Characteristic behaviors define each of the steps in this continuum. For example, in clinical problem solving the novice is rule-driven and in conjunction with lack of clinical experience relies solely on analytic reasoning, using every piece of information whether or not it is relevant. The advanced beginner is learning to sort the relevant from the irrelevant and even which rules are important. He now has some experience and can use pattern recognition, in addition to analytic reasoning, to solve problems. As one continues to progress from competent through master, ongoing experience allows for what appears to be almost intuitive problem solving and ultimately a practical wisdom a the mastery level. Likewise the Reporter-Interpreter-Manager-Educator (RIME) model of cognitive and skill development, has been found to be helpful in providing narrative anchors that enhance the reliability of assessment.[12] A less well-known developmental model in medical education is that of *identity development*.[13] This model is particularly helpful in providing insight into the developmental progression of professionalism and communication skills. These models as well as literature that elucidates the development of each of the behavioral elements within a given sub-competency are being used to inform the "Pediatrics Milestones Project", a joint initiative of the ACGME and ABP to further define and refine the ACGME competencies within the context of our specialty.[14-16]

Balmer et al[17] directly observed 143 hours of routine activities on an pediatric inpatient unit, and demonstrated that while the ACGME competencies were not explicitly named, all of the clinical discussions could be mapped to the competencies, suggesting that these competencies are implicit in all clinical work. Recent literature that integrates the six individual competencies and reframes them in the clinical context of the professional activities of the specialty provides a meaningful way of both teaching and assessing overall clinical competence. Ten Cate and colleagues suggest that competence in each specialty can be defined by 50-100 "*entrustable professional activities*" (EPAs).[1] Each EPA represents an integration of the competencies within a clinical context. An example of an EPA in pediatrics would be care of the nor-

mal newborn. In order to demonstrate competence in performing this activity one must have knowledge of maternal conditions affecting the infant, be able to perform a thorough physical examination with attention to congenital abnormalities, educate the mother about caring for her newborn using language that is understandable to her and respectful of her cultural background and its child rearing practices, and also provide continuity in transferring care from the hospital to the community provider. Close examination of the elements of this activity shows that it easily maps to the six ACGME competencies, to specific sub-competencies and then *milestones*. The word "entrustable" identifies an essential element of this conceptual framework, that is, the relationship between the faculty supervisor and the trainee that allows the faculty member to determine when the trainee is competent to perform the professional activity without direct supervision and can therefore be entrusted to do so. The act of entrustment is intuitive to clinician-educators and thereby provides a meaningful way of assessing competence using needed degree and type of supervision as the gradient. Work by Kennedy and colleagues around levels of supervision shows much promise in contributing a practical and meaningful strategy that can be used in conjunction with other methods to assess competence.[18]

The debate about assessing individual elements of competencies or integrating them should focus on the balance rather than the dichotomy. Only then will we be able to integrate the competencies into education and practice such that they become habits of care and, in turn, provide a comprehensive assessment of the learner that leads to meaningful feedback, improvement in care delivery, and the "good doctor" as the outcome of our efforts. Putting the competencies into the context of what we do as pediatricians makes the teaching and assessment of them much more meaningful.

Lessons Learned

- While tools that focus on specific components of competencies are important in identifying baseline skill sets, equally important is the assessment of the learner's ability to put these skills together to perform the professional activities expected of a pediatrician.
- In order for assessment to be meaningful and comprehensive, multiple methods, raters, and tools are necessary.
- Many factors affect assessment, such as the learning environment, training and calibration of raters, learner characteristics and competence, the quality of the assessment tools, the relationship between the faculty rater and the trainee and the availability of, interest in, and use of faculty development among the raters
- The road to competence is a developmental progression marked by the achievement of milestones along the way
- Longitudinal relationships between residents and faculty are a key element to making observations of resident performance meaningful and helpful for assessment.

References

1. Ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice. *Academic Medicine* 2007; 82:542-547.
2. van der Vleuten, CPM. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education* 1996; 1;41-67.
3. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical Education* 2005; 39:309-317.
4. Accreditation Council for Graduate Medical Education. Outcomes Project. [Internet] Available at: http://www.acgme.org/outcome. Accessed May 17, 2011.
5. Schuwirth LWT, van der Vleuten CPM. Changing education, changing assessment changing research? *Medical Education* 2004; 38:805-812.

6. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* 2004; 38: 327-333.
7. Holmboe ES. Direct observation by faculty. In Holmboe ES, Hawkins RE, editors. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia, PA: Mosby Inc.; 2008. p. 119-29.
8. Norman G, Bordage G, Page G, Keane D. How specific is case specificity? *Medical Education* 2006; 40: 618-623.
9. American Academy of Pediatrics, Committee on Children with Disabilities. Developmental surveillance and screening of infants and young children. *Pediatrics* 2001; 108(1):192—196
10. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Affairs* 2002; 21:103-111.
11. Carraccio CL. Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: Translating the Dreyfus developmental model to the learning of clinical skills. *Academic Medicine* 2008; 83:761-767.
12. Pangaro L. Investing in descriptive evaluation: a vision for the future of assessment. *Medical Teacher* 2000; 22:478-481.
13. Forsythe G. Identity development in professional education. *Academic Medicine* 2005; 80:S112-117.
14. Nasca T. Where will the ''Milestones'' take us? The next accreditation system. *ACGME Bulletin* September 2008; 3—5.
15. Hicks PJ, Schumacher DJ, Benson BJ, Burke AE, Englander R, Guralnick S, Ludwig S, Carraccio C. The Pediatrics milestones: Conceptual framework, guiding principles, and approach to development. *Journal of Graduate Medical Education* 2010; 2(3):410-418.
16. Hicks PJ, Englander R, Schumacher DJ, Burke A, Benson BJ, Guralnick S, Ludwig S, Carraccio C. Pediatrics Milestone Project: Next steps toward meaningful outcomes assessment. *Journal of Graduate Medical Education* 2010; 2(4): 577-584.
17. Balmer DF, Master CL, Richards B, Giordano AP. Implicit versus explicit curricula in Pediatrics: Is there a convergence? *Pediatrics* 2009; 24:e347-e354.
18. Kennedy TTJ, Lingard L, Baker GR, Kitchen L, Regehr G. Clinical oversight: conceptualizing the relationship between supervision and safety. *Journal of General Internal Medicine* 2007; 22:1080-1085.

Annotated Bibliography

Ten Cate O, Scheele F. Competency-based postgraduate training: Can we bridge the gap between theory and clinical practice. *Academic Medicine* 2007; 82:542-547.

The authors propose that the integration of the competencies in performing the routine activities of the profession is a meaningful focus for assessment. They go on to define "entrustable professional activities (EPAs)" as all the professional activities that a specific medical specialist must perform. A simple two-dimensional matrix demonstrates the alignment of EPAs with the ACGME competencies. The authors believe that 50-100 EPAs can define a training program of 5-6 years in duration. They provide a list of 8 conditions for defining EPAs. The EPAs are the targets for assessment. They are "units of work" that a supervisor can witness and attest to the ability of the learner to assume responsibility for in practice. "A statement of awarded responsibility (STAR)" can be awarded when the threshold for independent practice is reached. A new national OB/Gyn curriculum from the Netherlands uses the following framework for awarding STARS based on predetermined criteria as follows: "1) has knowledge, 2) may act under full supervision, 3) may act under moderate supervision, 4) may act independently (STAR) and 5) may act as a supervisor and instructor". The learner must be assessed in the context of delivering care and in order to engage clinicians in assessment one cannot detach assessment from the actual care delivery process/setting. They argue that "the supervisor's subjective but expert judgment is potentially a richer source of information than most other methods of assessment".

van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to pro-grammes. *Medical Education* **2005; 39:309-317.**

The authors discuss a conceptual model for assessment that relies on not just reliability and validity but educational impact, acceptability to stakeholders and investment in resources. Each element may receive different weight in different contexts. These weighted elements, in the aggregate, define the utility of the method. Other salient points are that improvement in reliability that was demonstrated with the OSCE resulted not from "standardization" but rather better sampling (larger samples across different patients and examiners). They argue for integration of educational components and a "whole task" approach to assessment. "Stacking" of components or sub-skills of competencies are less effective than methods in which different task components are presented and practiced in an integrated fashion. They warn that "at-omization my lead to trivialization and may threaten validity". It is important to use multiple types of as-sessment, including qualitative measures in order to create a meaningful whole. These last two points are critical for understanding how one makes assessment itself a meaningful learning tool.

van der Vleuten CPM. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education* **1996; 1:41-67.**

This is a helpful article for framing assessment and thinking about how one creates a system of assess-ment. It discusses fours classes of methods that attempt to measure different aspects of competence: mul-tiple-choice questions, written simulations, learning process measures and live simulations. The author explores issues of reliability, validity, and most importantly a modern view of competence, as well as the utility of assessment methods and their implications for practice and research. The take home message from the modern view of competence emphasizes that the transition to expertise involves a transition from a primarily analytic to non-analytic ability to handle clinical encounters effectively and efficiently based on past experience but warns that this ability is relatively dependent on the specific situations and not easily transferred from one context to another. Utility of an assessment tool or system is the multipli-cative result of reliability, validity, educational impact, acceptability and cost. He makes the point that each is an essential element for consideration in developing assessment tools and systems and if one of the elements is assigned a theoretical score of zero, the utility is zero. Under the final section he gives sage "suggestions" including the following highlights: (1) for reliability, "wide sampling …is imperative to allow for stable and reproducible scores"; (2) for validity, he supports a greater emphasis on direct val-idation studies where validity is built into the tool or test through a careful definition of what is being as-sessed and how it is being assessed as these dictate the essence of what is being measured; and (3) as-sessment drives learning through its content, format, and the information given. He states, "Instead of a decision tool, assessment should also be a learning exercise… We would argue that educational impact is the heart of educational achievement testing: assessment should be part of the learning process in order to achieve educational objectives set out in the training program."

7. Patient Care

Ann Burke, MD

"The good physician treats the disease; the great physician treats the patient who has the disease."
> **- Sir William Osler**

The Competency Defined

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 demonstrate competence in the following elements of patient care:

- Gathering essential and accurate information about the patient
- Providing transfer of care that insures seamless transitions
- Interviewing patients/families about particulars of the medical condition for which they seek care, with specific attention to behavioral, psychosocial, environmental, and family unit correlates of disease
- Performing complete and accurate physical examinations
- Making informed diagnostic and therapeutic decisions
- Developing and carrying out management plans
- Prescribing and performing all medical procedures
- Counseling patients and families
- Providing effective health maintenance and anticipatory guidance
- Using information technology to optimize patient care[1]

Rationale

The patient care competency includes multiple components that overlap with other competencies. For example, "counseling patients and families" encompasses features of interpersonal and communication skills, as well as professionalism. This chapter explores tools that can be used to assess many distinctive elements of the patient care competency; other tools are covered in other chapters.

Competence is contextual.[2,3] That is, competence reflects the relationship between a resident's abilities and the tasks she performs in a specific situation. Patient care sub-competencies, such as clinical reasoning and history taking, have content-specific characteristics as well. As a result, patient care assessments of trainees are not necessarily generalizable from one case to another or from one context to another.[2,3]

The Accreditation Council for Graduate Medical Education's (ACGME) "Companion Document"[4] states, "For all elements of patient care, direct observation of the resident's skills is critically important." It also suggests supplementing direct observations with such methods as *chart-stimulated recall* (CSR), standardized patients, and simulations. The program director must apply assessment techniques that are both feasible and meaningful measurements of resident skill.

Goals

1. Be able to identify methods and strategies to assess patient care competency domains in a residency program. Gain familiarity with these methods' strengths and weaknesses.
2. Understand that the complex interplay of assessment methods for one competency may overlap with other competencies.
3. Understand the concept of the utility model; plan for programmatic assessment in a manner that is feasible and meaningful.
4. Understand that not all programs will assess patient care in the same manner and that it is important to align program goals and resources with assessment methods.

Case Example

You are a new program director. The former director did not leave you with many instructions. You are feeling disorganized and overwhelmed. Lately, you have been trying to figure out a strategy to assess patient care knowledge, skills, and attitudes of the trainees. Past practice had been evaluations of a global sort, with vague, broad questions, which were supposed to be completed by faculty at the end of each rotation. The global rating anchors were "exceeds expectations", "meets expectations", and "needs improvement" in various patient care domains; there were no descriptive anchors. You feel that the files of the residents are somewhat sparse. When files included evaluations, the evaluations generally recorded fives (out of five) and nonspecific comments like "pleasure to work with". You are interested in figuring out what tools and types of assessment methods are available and what their strengths are. You want to make fair and valid decisions based on accurate, reliable assessments of your residents.

Points for Consideration

What are the assessment methods for documenting Patient Care competence?

Two major approaches to assessing the patient care competency are *observations of clinical performance* and *performance tests.*

Observations of clinical performance occur in actual patient settings. They can range from informal "snapshot" observations of learners in clinical settings to complex and formal systems that include multiple raters providing assessment data about residents in several clinical settings over time.[5] These observations represent the highest level of assessment in Miller's Pyramid—the "Does" pinnacle—as illustrated in the figure below.[5,6] Observations of clinical performance sacrifice some standardization and control of the setting and situation of assessment in favor of unprompted *authenticity*. Examples include live performance assessment with direct observation, undercover simulated patients, and video recordings of trainees. Observation may also include global assessments of residents; such assessments can provide useful, meaningful, and reliable measures when the assessor has observed the trainee often enough, and long enough, and knows what to look for.

Performance test is a generic term used to describe a number of types of formal testing such as the *objective structured clinical examination* (OSCE), simulations, and *standardized patients*. Performance tests measure what trainees <u>can</u> do when they know they are being assessed. This corresponds to the second highest level on Miller's Pyramid, "Shows".
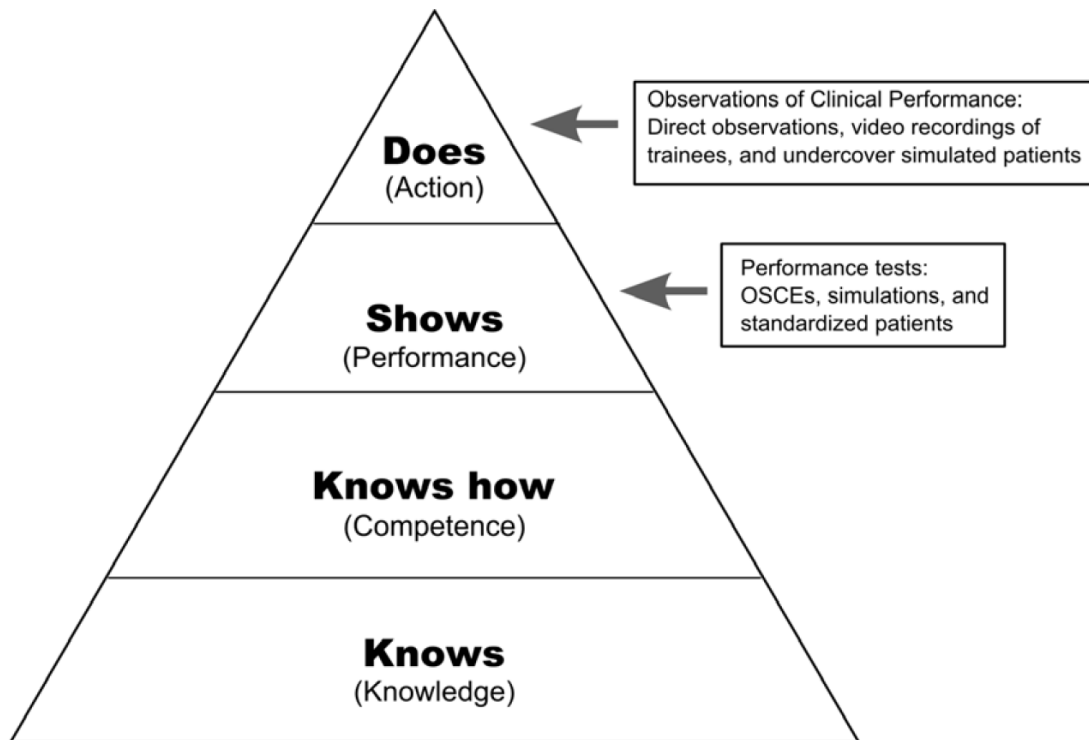
Miller's Pyramid. From Miller GE. 1990. The assesment of clinical skills/competence/ performance. Academic Medicine, 64(9):S63-7. Copyright © Association of American Medical Colleges. Reproduced with permission of Lippincott Williams & Wilkins.

In addition to direct observation, there are other approaches to assess elements of the patient care competency. These include chart-stimulated recall (CSR), chart reviews, and procedure logs; each of which can provide evidence about a trainee's knowledge and skill. A complete program evaluation system should include a number of these assessment methods, as well as direct observation. Various, specific tools can be used in all of these assessment methods, including global ratings and checklist evaluations. There are any number of tables and descriptions of various assessment tools in patient care,[3,5,7,8] and program directors should review the validity evidence supporting each assessment tool and method they propose to apply. The table on the following page summarizes some of the methods described in this chapter.

What kinds of assessment (and teaching) methods are classified as observations of clinical performance?

Direct Observation: There are multiple tools available for direct observation and assessment of clinical skills in trainees. A systematic review recently identified 55 such tools in the literature, but noted that validity evidence and description of educational outcomes are scarce.[8] Arguably the most studied tool is the Mini-CEX.[8,9] The Mini-CEX is an assessment in which the attending physician observes the trainee engaged in a patient encounter, performing skills such as the history and/or parts of the physical exam. This kind of structured observation with real patients and faculty observers, however, can have the same reliability as structured examinations using standardized patients.[10] Four Mini-CEX assessments in the same context are adequate to achieve sufficient reliability.[9] However, it may be difficult to get faculty members to accomplish that many even in one year.[5,9] Another disadvantage of the Mini-CEX is that the observations are task- and content-specific.

Method and Definition	Strengths	Limitations
Global Ratings Rating scales used to assess performance in authentic clinical settings based on multiple observations over time; often Likert-like scales.	Provides dynamic, formative feedback to learners, true-to-life observations. Able to assess integrative functioning.	Need extensive rater training, must have faculty with sufficient exposure to observe the learner on multiple occasions.
Checklist Evaluation Observational methods used to rate specific aspects of performance (behaviors) in clinical settings; typically yes/no items.	Allows for detailed feedback to trainee, helpful with technical skills assessment.	May be difficult to know how to score and weigh various items. Need expertise to develop checklists that have sufficient validity evidence.
OSCE Structured, standardized performance assessments, administered in sequential stations.	Allows for standardization across learners, control of the cases, and ability to provide complexity.	Logistics are complex, can be expensive, ethical and practical limitations to the use of children as standardized patients.
Simulations and Models Performance tests that attempt to model real life settings, with varying levels of fidelity.	Allows for standardization across learners, control of the cases, and ability to provide complexity.	Logistically complex, costs can be prohibitive, adequate sampling and ability to extrapolate to real-life are a concern.
Multi-Source Assessment (aka 360 Degree Evaluation) Forms or checklists that are completed by assessors with different perspectives: patients, families, attending, peers, nurses, hospital staff. May also include self-assessment.	Allows for *triangulation* and provides important formative feedback to learners.	Time intensive, may have high inter-rater variability, must train staff in rating scale/forms, many assessments needed to obtain reasonable reliability.
Standardized Patients Trained simulated patients that portray specific scenarios/cases and then rate the trainee's performance. May be "announced" (trainees know they are examining a standardized patient) or "unannounced" (trainees believe they are examining an actual patient; also called "undercover").	Allows for standardization and choice of cases, can be well-controlled which increases validity evidence.	Expensive, ethical and practical limitations to the use of children as standardized patients.

Strengths and limitations of methods for assessing patient care. Adapted from Downing and Yudkowsky[5] and ACGME/ABMS tables[7]

Direct observation without a standardized checklist or tool is also helpful for providing formative feedback, especially if the learner and observing attending physician work closely together in an apprenticeship-type model. However, with this method there is no "gold standard" for the various aspects of the exam or procedure, and ratings of different observers may vary considerably based on personal preferences of style of the various components of patient care.[11,12] Multisource assessment can be helpful to resident learners. The perspective of a peer may be different from that of an attending, possibly due to time spent accomplishing direct observation of patient care activities.

Videotaped Clinical Encounters: This type of clinical observation can offer a rich learning experience for trainees, but has many practical challenges. Videotaping the encounters requires the informed consent of patients and an exam room with video recording equipment, which is costly. Reviewing videotaped encounters is time consuming. On the other hand, the powerful feedback that can be provided to learners by viewing their own performance with real patients is worth these costs to many programs.[5]

<u>How else can I improve the validity and utility of the results of my observational assessments?</u>

Practical recommendations for enhancing the meaningfulness and usefulness of observational assessments of clinical performance include the following:[5, 12]

- Assessments should cover a broad range of clinical situations and procedures if you seek to draw conclusions about the resident's overall patient care competence.
- Formative assessments for teaching and learning should be separate from assessments performed for the purpose of learner promotion.
- Residents should be observed by multiple observers to reduce the effects of inter-observer differences.
- Rating tools and checklists should be short and focused and utilize descriptive anchors.
- Educate raters to make sure they are familiar with the assessment tools.
- Ensure that raters observe and rate specific resident behaviors/performance.
- Provide sufficient time for the observation session, such that assessments are thoughtful and candid, not rushed.
- Observational data should be recorded directly after the observation to prevent bias or change in scoring due to forgetting important elements or misplacing information.
- Give faculty raters feedback about their severity and leniency to prevent them from becoming more strict or lenient than their peers.
- Supplement traditional observational assessments with standardized clinical encounters (simulated patients) and skills training.
- Acknowledge the limitations of observational assessment methods even while continually working to improve their quality.

<u>What are examples of *performance tests*?</u>

Performance tests provide an opportunity for residents to "show how" they respond to complex cases and challenges, while controlling many factors of the case and context. These types of tests thus allow for standardization across residents. Performance tests can, however, be logistically complex, expensive and time-consuming. There may also be some difficulty in realistically modeling clinical situations, especially in pediatrics.[13]

Standardized Patients*:* A standardized patient (SP) is a person—often a professional actor—who is trained to act like a patient or the parent of a patient. This portrayal requires a highly detailed script and rigorous training. In some instances, like an OSCE (discussed in greater detail below), the SP must repeat the same performance for each student assigned to the station, making consistency in performance a critical element. In pediatrics, SPs are difficult, if not impossible, to use for some scenarios. For example, although it is possible for an actor to portray the parent of a newborn, it is impossible to portray the baby himself in a standardized manner. Some SPs are trained to provide feedback directly to the trainee; others are trained to rate the trainee's performance. SPs can be expensive and it is time-intensive both to write scripts and to train actors.[7,14]

The Objective Structured Clinical Examination (OSCE): An OSCE is an exam format that consists of a series of "stations" to test performance. Often, stations include the examination of standardized patients, but other simulations, including high fidelity simulations of procedures, writing clinical notes, and interpreting laboratory results may also be employed. First described in 1975,[15] OSCEs provide program directors and educators with a solution to the validity threats posed by *case specificity*— the finding that performance on one clinical case or station is often a poor predictor of performance on another. For ex-

ample, the ability to manage a three year-old patient with a pleural effusion does not predict the ability to provide anticipatory guidance for a six month-old, successfully complete a lumbar puncture, or conduct an appropriate physical exam on a teenager. A larger number of stations allows for better sampling of the patient care skills to be assessed, thereby improving the validity of the exam scores.[5, 11] Each OSCE station may require from five to thirty minutes, depending on the exam. Shorter stations are appropriate for discrete skills such as interpretation of an EKG or use of an otoscope to examine the ear. Longer stations are required for more complex cases to assess clinical reasoning or skill at counseling a patient.[3,5,7] As a rule of thumb, approximately ten to twelve stations may be required to achieve minimal generalizability.[5] Interestingly, it does not appear to matter if scoring in an SP station is done by the standardized patient or an examiner/observer.[14,15]

Simulations: Simulations include a wide array of tools ranging from static manikin heads for intubation to elaborate computer-based systems that are responsive to the trainee's actions. They can be used to assess single trainees or teams of residents. A systematic literature review of the uses and characteristics of high-fidelity simulation over 35 years emphasizes the importance of integrating simulations into an overall curriculum plan.[16]

Do I have to assess each element in the Patient Care competency and document "what the resident can do" on detailed checklists or will a "global" assessment of perceived competence at the end of the month suffice?

Global ratings based on observing samples of clinical performance of patient care elements are a primary means of assessing clinical competence.[5,12] Many programs use an "End of the Month" global rating form with items that rate performance as an integrated whole.[7,12] For example, faculty may be asked to determine whether, overall, a resident's performance was "unsatisfactory, marginal, good, very good, or outstanding". Raters judge general groupings of abilities, such as patient care, procedural skill, or professionalism, and the assessment is completed retrospectively based on general impressions over a period of time.[7]

Global ratings are subject to multiple threats to validity including limited direct observations of trainees, rater bias, inaccurate recall, and lack of specific clinical skills or tasks to rate.[12,17] Reliability of ratings can be improved by using descriptive anchors and providing sufficient faculty development and practice with the rating tool. Ideally, end of the month assessments should be supplemented with other assessment approaches, such as OSCEs, high-fidelity simulations, and directed clinical observations such as the Mini-CEX.

There is healthy debate about the value of tools according to their position on the spectrum from global ratings to detailed behavioral checklists. Neither global assessments nor checklist tools yield gold-standard judgments of resident ability.[2,3] Checklists seem onerous to many program directors, and because performance is often case-specific, the questions arises, "do I have to do detailed checklists on each and every sub-competency and every clinical situation in pediatrics to fairly and adequately assess each of my residents?" A global assessment is inherently more subjective, but may be sufficiently reliable if a faculty member spends significant time longitudinally with a trainee. Rinstead and colleagues conducted a study directly comparing the feasibility of checklists and global rating forms.[18] Utilizing 32 anesthesiology clinicians assessing anesthesiology residents in four simulated clinical scenarios, the authors compared global ratings with checklist scores. Clinicians felt that the checklist was significantly more appropriate than the global rating for this setting (assessment of specific skills and procedures). However, inter-rater agreement on pass or fail decisions was poor with both forms.

The clinical assessment of patient care, particularly using observational methods, depends on the availability of skilled, trained, and motivated faculty.[5] While many faculty may perceive themselves as expert

raters of resident performance, evidence suggests that greater seniority and clinical experience do not automatically make one a more reliable rater. Faculty require significant training and calibration to perform meaningful, reliable assessments of residents.[5,12] The balance for practical assessment probably lies in using multiple methods, knowing the shortcomings of both types of tools, applying principles of assessment to maximize validity evidence, and aligning the methods chosen with one's institutional assets.

<u>How do I begin to look at all of the different "tools" available and figure out how many to use and where to use them in my program to assess patient care?</u>

There are a myriad of assessment methods for the Patient Care competency, and each method has strengths and weaknesses. None provides a perfect assessment of a resident's progress. When evaluating which methods to use, it is helpful to consider the *utility model* (previously discussed in Chapter 1). This model encourages educators to consider five criteria: reliability, validity, impact on learning, acceptability/feasibility, and costs.[19] Depending on the purpose of the assessment, parameters may receive different weights. For example, high costs may be tolerated if the assessment has high stakes. However, a formative assessment that primarily provides feedback should be weighted more heavily on the impact on learning factor.[20] Feasibility can be challenging. For example, a 56-question checklist may not be adaptable to direct observations in the PICU setting. A recent study found that residents were infrequently observed by faculty performing basic patient care functions such as obtaining patient histories and performing physical exams.[21] Researchers in a study attempting to validate Mini-CEX scores found it difficult to get faculty to complete four Mini-CEXs on each intern throughout a full year.[9]

The complexity of patient care virtually mandates the use of multiple assessments.[5] *Triangulation*, or assessing and considering a resident from different perspectives and with various measures, can enhance validity. At the same time, the program director must avoid "assessment fatigue" or interfering with the care of patients. In each assessment, the items should be evaluated and mapped to determine whether and how they reflect the objective being assessed.[2] Therefore, no exact number or type of assessments is "correct". Program directors should adopt a utility-driven strategy for choosing and using meaningful "tests" to assess residents' competence in patient care.

Lessons Learned

- Patient care is case-specific.
- Multiple assessments are necessary to address the case specificity in the measurement of the patient care competency.
- OSCEs can be a good way to gather accurate information about patient care skills in a controlled setting.
- Global ratings are fraught with biases, particularly if faculty members have only limited contact with the learner. Direct observation of specific skills rated over time by the same rater can be reliable, however.
- Utility of the assessment methods for patient care must be considered. Even the most reliable tool is not useful if the faculty do not complete it.

References

1. Accreditation Council for Graduate Medical Education. Pediatric Common Program Requirements 2007. [Internet]
http://www.acgme.org/acWebsite/downloads/RRC_progReq/320_pediatrics_07012007.pdf. Accessed September 15, 2010

2. Wass V, van der Vleuten C, Shartzer J, Jones R. Assessment of clinical competence. *The Lancet* 2001; 357:945-949.

3. Epstein RM. Assessment in Medical Education. *New England Journal of Medicine* 2007; 356(4):387-396.

4. Accreditation Council for Graduate Medical Education. Pediatric Common Program Requirements, Companion Document 2007. [Internet]
http://www.acgme.org/acWebsite/downloads/RRC_progReq/320_pediatrics_core_companion.pdf
Accessed September 25, 2010

5. Downing S, Yudkowsky R, editors. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009.

6. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Medical Education* 2004; 38:327-333.

7. Accreditation Council for Graduate Medical Education, American Board of Medical Specialties. Toolbox of assessment methods. [Internet] 2000.
http://www.acgme.org/Outcome/assess/Toolbox.pdf . Accessed August 25, 2010

8. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees, a systematic review. *JAMA* 2009; 302(12):1316-1326.

9. Norcini JJ, Blank LL, Duffy FD, Fortna GS The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine* 2003; 138(6):476-481.

10. van der Vleuten CP, Norman GR, DeGraaf E. Pitfalls in pursuit of objectivity: issues of reliability. *Medical Education* 1991; 25:110-118.

11. Downing SM. Reliability: On the reproducibility of assessment data. *Medical Education* 2004; 38:1006-1012.

12. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical competence ratings. *Teaching and Learning in Medicine* 2003; 15:270-292.

13. Lane LJ, Ziv A, Boulet JR. A pediatric clinical skills assessment using children as standardized patients. Arch *Pediatric and Adolescent Medicine* 1999; 153:637-644.

14. van der Vleuten CM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine* 1990; 2:58-76.

15. Harden R, Stevensen M, Downie W, Wilson M. Assessment of clinical competence using objective structured examinations. *British Medical Journal* 1975; 1:447-451.

16. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher* 2005; 27:10-28.

17. Noel GL, Herbers JEJ, Caplow MP, Cooper MP, Pangaro LN, Harvery J. How well do internal medicine faculty members evaluate the clinical skills of residents? *Annals of Internal Medicine* 1992; 117:757-765.

18. Rinstead C, Ostergaard D et al. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. *Medical Teacher* 2003; 25(6):654-658.

19. van der Vleuten CPM. The assessment of professional competence: developments, research, and practical implications. *Advances in Health Science Education* 1996; 1:41-67.

20. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Medical Education* 2005; 39:309-317.

21. Holmboe ES. Faculty and the observation of trainees' clinical skills. *Academic Medicine* 2004; 79:16-22.

Annotated Bibliography

Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine* **2003; 15 (4):270-292.**

This comprehensive review describes the cognitive, social, and environmental factors that contribute unwanted sources of variation in scores in clinical performance assessments. The authors describe the various contexts in which performance assessments are carried out and review the evidence that bias is intrinsic to that process. They describe the available evidence for mechanisms that reduce bias, and they conclude by extrapolating sixteen recommended strategies for improving clinical practice assessments based on studies in both medical and non-medical contexts.

Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *The Lancet* **2001; 357 (4):945-949.**

This article discusses assessment of clinical competence which includes communication, interpersonal skills, and all of the other competencies, along with the patient care competency. Topics such as validity, reliability, Miller's Pyramid, blueprinting, and standard setting are described. Methods described include: multiple choice questions, short essays, OSCEs, oral cases and "long cases" (a method used in England). As the article summarizes, "Assessment at the apex of Miller's pyramid, the 'does', is the international challenge of the century for all involved in clinical competence testing. The development of reliable measurements of student performance with predictive validity of subsequent clinical competencies and a simultaneous educational role is a gold standard yet to be achieved."

Epstein RM. Assessment in medical education. *New England Journal of Medicine* **2007; 356(4):387-396.**

This is a foundational article for understanding assessment tools in medical education and would be valuable reading for every program director. The author provides a clear and easily approachable conceptual framework for many of the most commonly used assessment methods and the strengths and weaknesses of each. In addition to discussion of the most common written examination methods, the author also considers direct observation of live clinical encounters, clinical simulations with standardized patients and high fidelity mannequins, multi-source assessments and portfolios. The article highlights critical features of competence: (1) it is contextual. That is to say competence varies given different practice settings, disease prevalence, the nature of the patient's presenting symptoms, etc.; (2) aspects are content specific, and (3) the nature of acquisition of competence is developmental. Table 2 presents a wonderful summary of principles of assessment (goals, what to assess, how to assess, and cautions). The bibliography is an excellent resource. A must read for program directors.

8. Medical Knowledge

Richard Shugerman, MD

"In the end we can never be given knowledge by others; we can only be stimulated. We must develop our own knowledge."—Charles T. Tart

The Competency Defined

Residents must demonstrate knowledge of established and evolving biomedical, clinical, epidemiological and social-behavioral sciences, as well as the application of this knowledge to patient care. Residents must demonstrate sufficient knowledge of the basic and clinically supportive sciences appropriate to pediatrics. [1]

Rationale

While there has been considerable discussion in the medical literature about how to evaluate a physician's overall competence,[2] there has been considerably less discussion of how best to evaluate the individual competency of medical knowledge in resident physicians. Shortly after the introduction of the Accreditation Council for Graduate Medical Education (ACGME) general competencies, the educational literature was inundated with studies and commentaries on the appropriate methods for assessing new, unfamiliar, complex general competencies such as practice-based learning and improvement, systems-based practice, and professionalism. Medical knowledge, however, was either not mentioned or only cursorily discussed. In fact, medical knowledge has been assumed to be the one competency that program directors and faculty are comfortable assessing[3] and medical knowledge has been identified as the one competency that program directors feel most capable of remediating when specific deficits are detected.[4]

Much of the comfort with the assessment of medical knowledge likely stems from the availability of the *In-Training Examination* (ITE) which most ACGME accredited specialties employ. The ITE of the American Board of Pediatrics (ABP) has been available to assess individual resident knowledge in pediatric training programs since 1971. There is considerable evidence regarding the validity of ITE scores for predicting subsequent performance on the ABP General Certifying Exam.[5] For many pediatric residency programs, the ITE has become the gold standard for the assessment of medical knowledge among trainees.

Despite its ubiquity, the use of the ITE as the sole assessment of medical knowledge is inadequate. ITE performance is affected by the resident's general test-taking skills, a concern frequently raised by program directors when discussing residents with otherwise solid or outstanding performance evaluations who perform poorly on the ITE. Relying solely on the ITE for the assessment of medical knowledge is also inadequate because of the variability in the degree to which different residents prepare for the exam and the varying conditions under which different residencies administer the test. Fortunately, there are additional tools that can be utilized for the assessment of medical knowledge in pediatric residents. The strengths and weaknesses of the ITE and these additional tools for assessment of medical knowledge are the focus of this chapter.

Goals

1. Become familiar with a variety of tools for the assessment of medical knowledge in pediatric trainees and the strengths and weaknesses of each.

2. Consider the relationship of medical knowledge to the other core competencies and the manner in which assessment of one affects assessment of the others.

Case Example

The results of the In-Training Examination have just come in and you are reviewing individual scores for your trainees. As you move through the list of the PGY-2 class, you are struck by the surprisingly poor performance of a resident who you have gotten to know quite well. You have received several unsolicited emails and letters of praise for this resident from families, nurses and members of the faculty. You have just completed an inpatient rotation on which this resident was the senior for your team and you could easily identify the qualities of leadership, interpersonal communication skills, and professionalism that led to these accolades. While her fund of knowledge was never particularly remarkable in either a positive or negative way, she certainly seemed to have adequate knowledge to deliver excellent patient care and to lead your team effectively. You even made a note to yourself for her final evaluation that she had obviously been reading about her patients and that she demonstrated an appropriate knowledge as she taught during rounds.

You have been considering this resident as a potential Chief Resident if she maintains this level of clinical performance. You send a brief email to her to ask that she come and discuss the ITE results with you. She replies that she was hoping to sit down and discuss this matter with you as well, because her first year score was also quite low. Even though there was an increase from her first to second year scores, her score was still considerably below the mean for her level of training. Her email goes on to say that she has "always been horrible" at standardized tests and she is quite concerned about failing the Boards at the conclusion of residency.

Points for Consideration

What can you say to this resident about the validity of her score on the ITE as a measure of her medical knowledge?

The website of the ABP describes the purpose of the ITE as threefold: "to enable residents to assess strengths and weaknesses in general pediatric knowledge at the time of the examination; to assess their progress from year to year; and to compare their performance with national peer groups". The ABP's Program Director's Guidebook explains further: "The ACGME requires assessment of medical knowledge, and the ITE provides an ideal opportunity for a standardized annual assessment of medical knowledge for each resident."[6]

The ITE is composed of single-best-answer *multiple-choice questions* (MCQs). Single-best-answer MCQs are one of the most common forms of *selected-response formats* in all of written testing. A selected-response format is one in which the examinee is asked to select a response as opposed to a format such as essay or short answer in which examinees are asked to generate a response on their own. Selected-response questions in general, and single-best-answer MCQs in particular, have been intensively studied and have many advantages for the assessment of cognitive knowledge. According to Downing, "These objective written forms are efficiently computer scored, have very high agreement among content experts on the correctness of the keyed answer, have very strong and desirable educational measurement properties with an extensive research base and are typically easily defended if their construction has been carefully and systematically carried out."[7]

Many of the MCQs on the ITE are patient-based items that include laboratory and diagnostic findings in an effort to assess higher-order cognitive abilities required for clinical decision-making. There is consid-

erable support in the medical education literature for the validity of responses to "context-rich" multiple choice questions when the intent of the examination is to test the application of knowledge to clinical care.[8,9] There has been a tremendous amount of research into the most effective methods for writing MCQs and there is a great deal of agreement on the most important item writing guidelines.[10]

However, there are also important limitations to the MCQ format. One of the principal threats to the validity of MCQ scores as a measure of medical knowledge is the *cueing effect* in MCQ items—the potential for the examinee to recognize the correct answer in a list of possible answers when he would not have been able to provide the correct answer on his own had he been tested with an open-ended question. Most studies comparing multiple choice and open-ended questions in parallel formats have found that examinees score higher on multiple choice questions—an effect known as positive cueing.[11-13] A 1996 study by Schuwirth and colleagues demonstrated that MCQs can also exhibit negative cueing, in which the examinee would have generated the correct answer if asked an open-ended question but on the MCQ, he chooses the wrong answer due to the presence of plausible distracters in the list of possible choices. The authors concluded that the overall effect of cueing in MCQs is quite significant, occurring either positively or negatively for approximately 20% of questions.[14]

Concerns have also been raised regarding the effectiveness of testing medical knowledge with MCQs because MCQs do not replicate the real practice of clinical medicine. To paraphrase Downing, patients don't generally present to their physicians carrying a list of possible diagnoses from which the doctor must choose the single best answer.[7] Additional limitations of the MCQ format include the possibility of ambiguously written test questions, an examinee's opportunity to randomly guess the correct answer, and questionable effectiveness in measuring higher-order thinking. Most can be overcome with appropriate question writing techniques and with sufficient testing of questions in the field.[10]

Despite their limitations, single-best-answer, context-rich MCQs are the principal testing element of the ITE and also of the General Certifying Examination that all pediatric residency graduates must pass to obtain board certification. It is likely helpful for your resident to know both sides of the coin—that this testing method does have limitations as a measure of her pediatric knowledge but she will need to develop sufficient mastery of the format to reach her professional goal. For this particular resident, the result of the ITE should be an important element in a formative assessment providing clear guidance as to how much she will need to improve to pass the high stakes summative assessment of the General Certifying Examination.

Are there other written testing formats that might provide a different assessment of this resident's medical knowledge?

Aside from MCQs, other well-known written testing techniques include matching, true/false, fill-in-the-blanks, and essays. Essay exams are generally thought to assess higher order cognitive skills than MCQs[15] but the scoring of such exams is extremely expensive and very difficult to accomplish in a reliable manner. *Key feature testing*, *script concordance testing* and *computer-based testing* with case scenarios have also received considerable attention in large scale standardized settings. Each of these formats has been proposed to more accurately assess an examinee's diagnostic reasoning than MCQs and to do so in a more reliable manner than the essay format.

Key feature testing focuses on critical decision making in clinical settings. It is based on the concept that in any clinical encounter there are a small number of essential decisions that form the *key steps* or *key features* in the successful resolution of the problem.[16] By focusing questions on these key steps in clinical decision making, key feature testing was designed to build on the strengths and minimize the drawbacks of the lengthy and complex *patient management problems* (PMP's) that were popularized in clinical testing of knowledge in the early 1980's but lost favor due to concerns about low reliability.

Key feature testing is thought to assess real-world practice encounters more effectively than can be accomplished with single answer MCQs because it allows for more than one correct answer for any given clinical setting. Key feature testing has been used extensively by the Medical Council of Canada and the Royal Australian College of General Practitioners. In the United States, the American College of Physicians has utilized key feature testing in its Medical Knowledge Self Assessment Program.[17]

Script concordance testing is also thought to provide a more real world assessment of knowledge than single-best-answer MCQs. Script concordance questions attempt to measure the organization of clinical knowledge in the mind of the examinee and have been shown to be good predictors of clinical reasoning skills on oral exams.[18,19] Script concordance questions present an examinee with a clinical scenario and provide new elements of information in a stepwise fashion. Grading of the question is accomplished by comparing the concordance of the responses of the examinee with those of a panel of experts presented with the identical scenario.[20]

Computer-based testing is increasingly being used for high stakes examinations such as the USMLE. Computer-based testing offers several advantages over written paper tests—most importantly, a greater sense of authenticity and the ability to use multimedia questions. Computer-based testing often utilizes patient management problems (PMP) to assess problem solving skills. PMP's usually begin with a patient's presenting complaint. The examinee is asked to select appropriate items of history, examination, and investigation before making a diagnosis and outlining a management plan. Because a large number of PMP's are usually required to reliably test a candidate's problem solving ability, computer-based testing is an ideal format for these kinds of questions. On the other hand, computer-based testing can be expensive and the questions tend to be more "recallable" to trainees. This can impact examination security if individuals share notes about their examinations.[21]

There is little information in the literature about the use of any of these tools in standardized testing in pediatric residencies. Future availability of these assessment methods may provide program directors with alternative tools to assess medical knowledge in residents challenged by the MCQ format.

Beyond written testing, what other methods might you and your resident use to assess the adequacy of her medical knowledge?

Direct observation reported on a global assessment form is one of the most common methods in pediatric residencies for the assessment of medical knowledge. Although highly regarded by faculty members as a reliable and valid assessment technique—"Look, I know a smart resident when I see one"—the limited literature comparing faculty assessment of medical knowledge on global rating scales with an assessment of medical knowledge from the results of the ITE has shown little to no correlation. In one study, attending evaluations of medical knowledge from global assessment forms were compared with results of the ITE for 44 anesthesia residents.[22] Although no correlation was found between these two methods of assessing medical knowledge, strong correlations were found among faculty assessments of interpersonal skills, professionalism, and medical knowledge. The results raise questions about the influence of measures of social intelligence on the assessment of other competencies. The authors conclude that "residents deemed easy to work with may have a 'halo effect', which can alter assessments of medical knowledge". In the only other published study that directly addresses the reliability of direct observation as a tool for the assessment of medical knowledge, faculty in an Internal Medicine residency were asked to predict the tertile into which each of their 35 residents would score on the ITE. The overall accuracy of faculty prediction was 50% regardless of the amount of time that the faculty member had spent directly observing the residents' performance. Faculty tended to overestimate rather than underestimate the residents' scores.[23]

Chart-stimulated recall (CSR) is cited in the Toolbox of Assessment Methods on the ACGME website as one of the most desirable methods available for the assessment of medical knowledge. CSR exercises run the gamut from a formal standardized oral examination to an informal debriefing exercise at the end of a clinic session. In either setting, the goal of the CSR is to provide a window into the resident's thought processes and application of knowledge during an actual clinical encounter. In the more formal application, a trained physician examiner rates the resident using a well-established protocol and scoring procedure and the entire process can take as long as an hour to complete. As an informal debriefing tool after clinic, the CSR is often used by faculty members to assess a resident's clinical reasoning skills and to probe for evidence of limited knowledge base or premature diagnostic closure. If limited to a small number of patient encounters, the process may take less than 20 minutes but the reliability of the tool can be significantly diminished in such brief assessments.

<u>In the face of her disappointing ITE result, how can you help this resident to maintain the self-confidence that her recent outstanding clinical performance warrants?</u>

In counseling this resident about her performance, it is important to help her recognize the goals as well as the limits of our assessment techniques. It is generally agreed that one of the major goals for assessing medical knowledge is to provide direction and motivation for future learning. It should be relatively easy for this resident to understand that she has work to do to find a way to demonstrate in a standardized, written testing format the knowledge that she displays so easily in the clinical setting. While beyond the scope of this chapter, initial steps in remediating her performance might include a review of her previous history with standardized testing such as the MCAT and USMLE. If her pattern of poor test results is consistently borne out, she may well benefit from an evaluation with a testing specialist who could counsel her in both test preparation and test taking skills.

As important as setting goals and providing tools for improvement, reviewing the limitations of standardized testing with this resident is essential in maintaining her self-confidence and continued successful performance in the clinical setting. In his 2007 review of assessment techniques, Epstein describes the psychometric challenges in assessing clinical expertise:

> "Expertise is characterized by unique, elaborated, and well-organized bodies of knowledge that are often revealed only when they are triggered by characteristic clinical patterns. Thus experts who are unable to access their knowledge in artificial testing situations but who make sound judgments in practice may do poorly on some tests that are designed to assess communication, knowledge, or reasoning."[24]

By reminding this resident that there are significant limitations to all of our assessment techniques and that no single test provides an unbiased assessment of her medical knowledge, you may help to preserve the confidence and the determination she will require to continue working to improve her score.

Lessons Learned

- The single best answer MCQ using patient-based formats as employed on the ITE is a well-studied and highly objective tool to test the application of medical knowledge in clinical decision making. The cueing effect is one of the major threats to the validity for this assessment method.
- The results of the ITE are an important indication as to the likelihood of success on the General Pediatrics Certifying Examination.
- Direct observation in clinical care and chart-stimulated recall are two additional assessment tools that should be combined with the ITE in an overall evaluation of medical knowledge but both may be affected by interpersonal and communication skills.

References

1. Accreditation Council for Graduate Medical Education. Revised Common Program Requirements (Pediatrics). Revision of July 1, 2007. [Internet] *http://www.acgme.org/acWebsite/downloads/RRC_progReq/320_pediatrics_07012007.pdf. Accessed May 15, 2011.*

2. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002; 287:226-35.

3. Heard JK, Allen RM, Clardy J. Assessing the needs of residency program directors to meet the ACGME general competencies. *Academic Medicine* 2002; 77:750.

4. Torbeck L, Canal DF. Remediation practices for surgery residents. *American Journal of Surgery* 2009; 197:397-402.

5. Althouse LA, McGuinness GA. The in-training examination: an analysis of its predictive value on performance on the general pediatrics certification examination. *Journal of Pediatrics* 2008; 153(3):425-8.

6. American Board of Pediatrics. General Pediatrics Program Director's Guide to the ABP. 2009. [Internet] *https://www.abp.org/abpwebsite/publicat/pdguide09.pdf. Accessed May 15, 2011.*

7. Downing SM. Assessment of knowledge with written test forms. In Norman GR, Van der Vleuten CPM, Newble DI, editors. *International Handbook for Research in Medical Education*. Dordrecht, The Netherlands, Kluwer Academic Publishers; 2002. p. 647-672.

8. Schuwirth LWT, Verheggen MM, van der Vluten CPM, Boshuizen HPA, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? *Medical Education* 2001; 35:348-356

9. Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Medical Education* 2004; 38:974-979.

10. Haladyna TM, Downing SM, Rodriquez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education* 2002; 15:309-334.

11. Newble DI, Baxter A, Elmslie RG. A comparison of multiple choice and free response tests in examinations of clinical competence. *Medical Education* 1979; 13:263-268.

12. Case SM, Swanson DB. Extended-matching items: a practical alternative to free-response questions. *Teaching and Learning in Medicine* 1993; 5:107-115.

13. Veloski JJ, Rabinowitz HK, Robeson MR. A solution to the cueing effects of multiple-choice questions. *Medical Education* 1993; 27:371-375.

14. Schuwirth LWT, van der Vleuten CPM, Donkers HHLM. A closer look at cueing effects in multiple-choice questions. *Medical Education* 1996; 30:44-49.

15. Bennett, RE. On the meanings of constructed response. In Bennett RE, Wald WC, editors. *Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment*. Lawrence Erlbaum Associates; 1993. p. 1-28.

16. Pate G, Bordage G. The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Academic Medicine* 1995; 70:104-120.

17. Farmer EA, Page G. A practical guide to assessing clinical decision-making skills using the key features approach. *Medical Education* 2005; 39:1188-94.

18. Brailovsky C, Charlin B, Beausoleil S. Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency. *Medical Education* 2001; 35:430-6.

19. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten CPM. The script concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine* 2002; 12:189-195.

20. Charlin B, van der Vleuten CPM. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Evaluation & the Health Professions* 2004; 27:304-319.

21. Cantillon P, Irish B, Sales D. Using computers for assessment in medicine. *British Medical Journal* 2004; 329:604-609.

22. Minhaj MM, Klafta JM, Tung A. Good doctor? Nice person? Correlations between ACGME competencies in an anesthesia training program. American Society of Anesthesiologists Annual Meeting, October 2007.
23. Hawkins RE, Sumption KF, Gaglione MM, Holmboe, ES. The In-Training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty predictions of resident performance. *American Journal of Medicine* 1999; 106:206 —210.
24. Epstein RM. Assessment in Medical Education. *New England Journal of Medicine* 2007; 356:387-396.

Annotated Bibliography

Downing SM. Assessment of knowledge with written test forms. In Norman GR, Van der Vleuten, CPM, Newble, DI, editors. *International Handbook for Research in Medical Education*. Dordrecht, The Netherlands, Kluwer Academic Publishers; 2002. p. 647-672

This chapter provides the reader with fundamental information on the use of written assessment formats to test cognitive knowledge. The author discusses the basic features of selected-response and constructed-response formats, the psychometric properties associated with each and the strengths and weaknesses of each testing approach. It is clearly written and easily approachable—an extremely useful chapter for anyone seeking to understand the theoretical basis for written examinations. The bibliography is extensive.

Schuwirth LWT, van der Vleuten CPM, Donkers HHLM. A closer look at cueing effects in multiple-choice questions. *Medical Education* 1996; 30:44-49

This study of the cueing effect in multiple choice testing is a must read for anyone interested in understanding the variability in individual test-taking skills. The authors sought to quantify the degree to which the cueing effect alters test results in either a positive or a negative direction and the degree to which expertise in a particular field alters the magnitude of the cueing effect. The authors employed one test containing 35 identical clinical cases given to 75 medical students, 25 residents and 25 experienced practicing physicians. The test was administered to all examinees with both an open-ended response and with a multiple choice response format. Across all examinees, the impact of positive cueing was about 14%, the impact of negative cueing was about 7% and the overall net effect was about 7%. Difficult questions resulted in more positive cueing, easier items showed more negative cueing and experience decreased but did not eliminate the effect altogether.

Cantillon P, Irish B, Sales D. Using computers for assessment in medicine. *British Medical Journal* 2004; 329:604-609.

This brief overview of computer based testing in medicine is a valuable summary of the history of the approach, the advantages and disadvantages of the various formats and essential questions to be answered before choosing to employ a computer based examination.

9. Practice-based Learning and Improvement

Patricia Hicks, MD

"If I had an hour to solve a problem and my life depended on the solution, I would spend the first 55 minutes determining the proper question to ask, for once I know the proper question, I could solve the problem in less than five minutes." - Albert Einstein

The Competency Defined

Residents must demonstrate the ability to investigate and evaluate their care of patients, to appraise and assimilate scientific evidence, and to continuously improve patient care based on constant self-evaluation and life-long learning. Residents are expected to develop skills and habits to be able to meet the following goals:

- identify strengths, deficiencies, and limits in one's knowledge and expertise;
- set learning and improvement goals;
- identify and perform appropriate learning activities;
- systematically analyze practice using quality improvement methods, and implement changes with the goal of practice improvement;
- incorporate formative evaluation feedback into daily practice;
- locate, appraise, and assimilate evidence from scientific studies related to their patients' health problems;
- use information technology to optimize learning; and,
- participate in the education of patients, families, students, residents and other health professionals. [1]

Rationale

Perhaps Albert Einstein could construct an approach to seeking resources, appraise the evidence, and then construct a solution to the problem in five minutes but most others would take considerably longer. Nevertheless, formulating questions aimed at closing gaps in one's knowledge through a reflective and iterative process should dominate the activities of practice-based (life-long) learning and improvement.

The Practice-based Learning and Improvement (PBLI) competency encompasses a complex set of sub-competencies. Based on the concepts of the continuous quality improvement approach.[2] PBLI has been described as the ability to execute the following series of steps in a continuous cycle: (1) determine improvement needs; (2) identify and apply an intervention; (3) measure the impact of the intervention and inform the next cycle.

According to the Accreditation Council for Graduate Medical Education (ACGME) Assessment Toolbox,[1,3] the most desirable tool to analyze a resident's own PBLI is a portfolio (see Chapter 5). However, program assessment of a resident's competence in PBLI has been studied within several disciplines using different approaches, giving program directors a number of lenses through which to view assessment of PBLI.[4,5]

Assessment of the wide range of PBLI subcompetencies requires a collection of assessment methods rather than a single approach. This chapter aims to explore some of the published reports of assessment strategies and tools and to prompt readers to consider what approaches they might choose to take.

Goals

1. Understand and describe how the critical aspects of PBLI can be taught and assessed in the context of a meaningful learning activity.
2. Identify methods which assess learner outcomes for achievement of PBLI, with a focus on the strengths and limitations of each.
3. Understand the complex interaction between the learner and the setting or context in which the assessment is occurring and the impact of this interaction on assessment of the learner.

Case Example

As program director, you are interested in developing residents' abilities to recognize and fill gaps in knowledge in order to enhance their clinical decision-making. In the spirit of quality improvement, you want them to not only identify gaps in their current knowledge and understanding of patients under their care, but to be able to seek resources to learn more about these patients, apply that new learning to their patient care, and teach others about their new knowledge. You develop a new curriculum focusing on a "Critically Appraised Topic" where residents identify new clinical questions prompted by patient encounters, seek evidence through literature and consultative resources to answer their questions, appraise the evidence obtained, and then synthesize and apply that new evidence to address the clinical question. This curriculum will culminate in a presentation to their peers. If the prompt for the topic chosen was a patient adverse event, near miss, or other safety-related prompt, the resident will be expected to make recommendations about system or practice changes or to identify areas where further study is needed.

You ask yourself, "How can I assess the competence of my residents' practice-based learning and improvement competence when using a Critically Appraised Topic project as well as other curriculum aimed at developing PBLI competence?"

Points for Consideration

How will you assess your resident's ability to identify gaps in their clinical knowledge or skills?

Assessing gaps and strengths in clinical knowledge or skills often occurs during the dynamic process of instructional methods that involve learner exploration and questioning. The nature of the types of questions and approach to exploration chosen by the learner often informs the teacher (and hopefully the learner herself) about the learner's level of understanding of the subject. Kolb[6] suggests that reflection on previous experiences helps us to formulate hypothetical questions (prompted from recognition and response to perceived gaps in knowledge, skills or attitudes) and engage in "active experimentation" (reading, applying various new strategies/approaches, etc.) which informs our learning going forward. Schön[7] describes a process whereby one reflects on a clinical issue either during or outside of the immediate clinical situation. Such reflection, which embraces uncertainty, conflict, and ambiguity, pushes the physician towards seeking new knowledge or skills in an attempt to understand and then incorporate this new learning into practice. It is the identification of specific deficiencies or limitations in knowledge, skills or attitudes that is critical. Resource-seeking and question-asking skills are then required to identify and specifically sort out strengths and deficiencies. The formation of questions enables a learner to more clearly de-

termine the difference between their current and ideal knowledge and skills, to develop a vision of competence, and to reflect on the forces encouraging and impeding change.[8]

Curriculum designed to develop critical thinking skills, combined with assessment of types of questions posed and knowledge identified can be useful to residents in determining gaps in their knowledge.[9] The table below lists a set of questions that can be employed to enhance critical thinking; these same questions generated by the resident-learner may indicate awareness of gaps or limitations in knowledge and/or skills.

Generic Question	Specific Thinking Skills Induced
What are the strengths and weaknesses of …?	Analysis / inferencing
What is the difference between … and …?	Compare—contrast
Explain why … (explain how …)?	Analysis
What would happen if …?	Prediction / hypothesizing
What is the nature of …?	Analysis
Why is … happening?	Analysis / inferencing
What is a new example of …?	Application
How could … be used to …?	Application
What are the implications of …?	Analysis / inferencing
What is … analogous to?	Identification and creation of analogies and metaphors
What do we already know about …?	Activation of prior knowledge
How does … affect …?	Analysis of relationships (cause-effect)

Adapted from King A, Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal* (v29n2, pp. 303-323). Copyright © 1992 by American Educational Research Association. Reprinted by permission of SAGE Publications.

Faculty involvement in the assessment of PBLI is critical since the nature of many of the judgments require an understanding of the specific clinical context. Dependence on faculty observation of trainees' clinical skills is fraught with difficulties, however, because of limited faculty time, inconsistency in faculty expectations, and brevity of interactions between faculty and residents. Sources of bias in clinical performance ratings[10] are many; identification of learner gaps is often inconsistently assessed. Some educational settings, such as morning report, offer an ideal setting for teaching and assessing practice based learning and improvement by offering a forum in which the diagnostic reasoning process is explained and explored, clarifying areas of understanding and areas where further knowledge is needed.

How will you assess your residents' ability to identify resources to address gaps in their knowledge and/or skills?

A resident's skill at identification of resources (human, peer-reviewed journal publications, or other evidence) and subsequent selection of appropriate evidence to address questions can be assessed by multiple methods.

Assessment of the quality of literature search strategies is often based on measuring both process and outcomes.[11-13] Process measures include use of Medical Subject Headings (MeSH) terms, Boolean operators such as "or" and "and", appropriate combination of search concepts, and filtering results by using search terms such as "therapy" or "prognosis" or "diagnosis". Outcome measures include search precision, recall, and efficiency. Precision is the positive predictive value of the search (how many items retrieved in

the search were actually relevant), recall is the sensitivity of the search (how many relevant items were found in the search relative to the total number of relevant items that could possibly be found), and efficiency measures how quickly a search with a given precision and recall is conducted.

Assessment of information-seeking from human sources (consultation) may be achieved by having the consultant provide feedback on how well the resident specified the reason for consultation or resource sought, the clinical context (what is known and what is needed to fill the knowledge or skill gap), and the request for new information and evidence. Consultant interactions can be further leveraged through assessment of resident teach-back of consultative information received.[14]

How will you assess the resident's *performance in synthesizing and applying* their newly discovered knowledge and/or skills?

Skills in synthesizing and applying newly discovered knowledge include (1) recognizing the patient problem or clinical question; (2) collecting, organizing, and synthesizing evidence that helps to answer the question; (3) critically appraising that evidence; (4) understanding the study results as they relate to the question; and (5) applying the evidence to one's practice.[15] These elements can be assessed by the learner and the mentor through a *critically-appraised topic* project. Self-assessment of these skills serves to prompt and inform the learner about the processes involved. For the mentor, reviewing the approach used and the quality of the products of each step helps to frame the assessment for the critically appraised topic.

How will you assess the resident's *performance in presenting* their newly discovered knowledge and/or skills?

Residents' synthesis of new knowledge or evidence can be assessed by scoring, on a well-designed tool, their choice of content, organization, and presentation to their peers. Several tools have been developed to assess content, organization, clarity and other presentation elements and skills.[16,17] Assessment by faculty, peers, and the resident themselves may be superior to evaluation by any one source alone.[18] The authors of these tools have suggested benchmarks representing a threshold of competence; no validity evidence was provided.[18]

In assessing your resident's performance in applying this new knowledge in the patient care setting, how will you sort out the difference between actual resident performance measures and institutional-system influences affecting performance outcomes demonstrated by the resident?

Assessment of an individual without simultaneous assessment of the system is incomplete. The institutional system issues that impact the learner's performance are constantly in flux. An ideal learner may not be able to perform optimally if the system or infrastructure is not supportive of such performance; excellent systems can be protective and assist learner performance. Distinguishing learner performance independent of the system's influence can be done with assessment of skills in a controlled objective structured clinical examination (OSCE). A pilot study reported high validity evidence for an OSCE to assess System-Based Practice (SBP; see Chapter 12) and PBLI.[19] Nine fellows in preventative medicine and endocrinology participated in an 8 station OSCE designed to assess competency in SBP and PBLI. A combination of written, standardized patient (SP) and simulation-based stations tested the knowledge, skills and attitudes, required for SBP and PBLI. Not all stations used SPs because some of the domains, such as the creation of quality measures and conducting a root cause analysis, require demonstration of knowledge and skills through written or graphical representation.[20] Whether results of performance on

such an OSCE can represent or predict performance in a changing healthcare environment is an important question because it is the real-world performance that impacts patient care.

Resident knowledge of practice-based learning and improvement learning elements involved in quality improvement (QI) projects can be systematically assessed. In internal medicine, the Quality Improvement Proposal Assessment Tool (QIPAT-7)[21] has been used as a measure of quality improvement learning that incorporates steps to bring about rapid-cycle improvement using a QI proposal.[22] Outcomes of such projects are greatly influenced by system factors. Assessment can examine educational and scholarly productivity and patient care outcomes after implementation of the proposed changes to strengthen validity evidence by demonstrating relationships among outcomes.[23]

What are other examples of professional activities that would be considered PBLI and how might these activities assist your ability to assess the resident's competence in PBLI?

Scholarly projects are optional in pediatric categorical training, although many programs require advocacy projects in addition to projects assigned as part of a systems-based practice improvement or PBLI curricular assignment.[24,25] A *case report poster* can be an appropriate resident-level scholarly project.[26,27] With proper mentorship, residents can achieve—and be assessed on—most, if not all, of the PBLI sub-competencies in constructing a case report poster.

PBLI work in the form of a systems *quality improvement project* is a common approach, because the ACGME competencies call for residents to be actively engaged in (SBP) improvement.[28] Assessment of resident achievement of PBLI in a SBP project based on the Plan-Do-Study-Act (PDSA) cycle was studied by Tomolo and colleagues.[29] They offer guidance for assessment while addressing the challenges of PDSA as a conceptual framework in the moving target of process improvement and resident education.

Lessons Learned

- The ACGME defines Practice-based Learning and Improvement (PBLI) as the resident's ability to investigate and evaluate his or her patient care practices, appraise and assimilate scientific evidence, and improve patient care practices. In other words PBLI is "how you get better" at medicine.[30]
- PBLI is an iterative and multi-faceted improvement process; as such, competence is best assessed using multiple methods.
- Assessment methods for PBLI and SBP overlap since both are grounded in continuous process improvement.
- Curricular activities such as preparation and presentation for journal club or a critically appraised topic are ideal for global and individual element assessment of PBLI.

References

1. Accreditation Council for Graduate Medical Education. *The ACGME Outcome Project: An Introduction*. [Internet] 2005. *http://www.acgme.org/outcome/*. Accessed May 18, 2011.
2. Ziegelstein RC, Fiebach NH. "The mirror" and "the village": A new method for teaching practice-based learning and improvement and systems based practice. *Academic Medicine.* 2004; 79:83-88.
3. Accreditation Council for Graduate Medical Education, American Board of Medical Specialties. *Toolbox of Assessment Methods.* [Internet] 2002. *http://www.acgme.org/outcome/assess/Toolbox.pdf*. Accessed May 18, 2011.

4. Langley GJ, Nolan KM, Nolan TW. *The Foundation of Improvement*. Silver Springs, MD: API Publishing; 1992.

5. Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing Practice-Based Learning and Improvement. *Teaching and Learning in Medicine.* 2004; 16(1):85-92.

6. Kolb D. *Experiential Learning as the Science of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall; 1984.

7. Schön D. *Educating the Reflective Practitioner*. San Francisco CA: Jossey-Bass Publishers; 1987.

8. Fox RD, Mazmanian PE, Putnam RW. *Changing and Learning in the Lives of Physicians*. New York, NY: Praeger Publishers; 1989.

9. King A. Designing the instructional process to enhance critical thinking across the curriculum. *Teaching of Psychology.* 1995; 22:13-17.

10. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine.* 2003; 15:270-292.

11. Voisin CE, de la Varre C, Whitener L, Gartlehner G. Strategies in assessing the need for updating evidence-based guidelines for six clinical topics: an exploration of two search methodologies. *Health Information and Libraries Journal.* 2008; 25:198-207.

12. Gruppen LD, Gurpreet KR, Arndt TS. A controlled comparison study of the efficacy of training medial students in evidence-based medicine literature searching skills. *Academic Medicine.* 2005; 80(10):940-944.

13. Nicholson LJ, Warde CM, Boker JR. Faculty training in evidence-based medicine: Improving evidence acquisition and critical appraisal. *Journal of Continuing Education in the Health Professions.* 2007; 27(1):28-33.

14. Silbert L, Lachkar A, Grise P. Communication between consultants and referring physicians: A qualitative study to define learning and assessment objectives in a specialty residency program. *Teaching and Learning in Medicine.* 2002; 14(1):15-19.

15. Hatala R, Keitz SA, Wilson MC, Guyatt G. Beyond Journal Clubs: Moving toward an integrated evidence-based medicine curriculum. *Journal of General Internal Medicine.* 2006; 21:538-541.

16. Jahangiri L, Mucciolo T. Presentation Skills Assessment Tools. *MedEdPORTAL.* 2006;7930. http://services.aamc.org/30/mededportal/servlet/s/segment/mededportal/find_resources/browse/?subid=7930. Accessed July 1, 2010.

17. Musial JL, Rubinfeld IS, Parker AO, Reichkert CA, Adams SA, Roa S, Shepard AD. Developing a Scoring Rubric for Resident Research Presentations: A Pilot Study. *Journal of Surgical Research.* 2007; 142:304-307.

18. Jahangiri L, Mucciolo TW, Choi M, Spielman AI. Assessment of Teaching Effectiveness in U.S. Dental Schools and the Value of Triangulation. *Journal of Dental Education.* 2008; 72(6):707-718.

19. Varkey P, Natt N, Lesnick T, Downing S, Yudkowsky R. Validity evidence for an OSCE to assess competency in Systems-based Practice and Practice-based Learning and Improvement: A preliminary investigation. *Academic Medicine.* 2008; 83(8):775-780.

20. Searle J. Defining competency - The role of standard setting. *Medical Education.* 2000; 34:363-366.

21. Leenstra JL, Beckman TJ, Reed DA, Mundell WC, Thomas KG, Krajicek BJ, Cha SS, Kolars, JC, McDonald FS. Validation of a method for assessing resident physicians' quality improvement proposals. *Journal of General Internal Medicine.* 2007; 22(9):1330-1334.

22. Morrison LJ, Headrick LA, Ogrine G, Foster T. The quality improvement knowledge application tool: An instrument to assess knowledge application in practice-based learning and improvement. *Society of General Internal Medicine Meeting.* Vancouver, BC; 2003.

23. Messick S. Validity. In: Linn RL, editor. *Educational Measurement*. 3rd ed. Phoenix, AZ: Oryx Press; 1993.

24. Summers RL, Fish S, Blanda M, Terndrup T. Assessment of the "Scholarly Project" requirement for emergency medicine residents: Report of the SAEM Research Directors' Workshop. *Academic Emergency Medicine.* 1999; 6:1160-1165.
25. Summers RL, Woodward LH, Sanders DY, Galli RL. Research curriculum for residents based on the structure of the scientific method. *Medical Teacher.* 1998; 20:36-37.
26. Willett LL, Paranjape A, Estrada C. Identifying key components for an effective case report poster: An observational study. *Journal of General Internal Medicine.* 2008; 24(3):393-397.
27. Carroll AE, Sox CM, Tarini BA, Ringold S, Christakis DA. Does presentation format at the Pediatric Academic Societies' Annual Meeting predict subsequent publication? *Pediatrics.* 2003; 12:1238-1241.
28. Leach DC. Evaluation of competency: An ACGME perspective. *American Journal of Physical Medicine and Rehabilitation.* 2002; 79:487-489.
29. Tomolo AM, Lawrence RH, Aron DC. A case study of translating ACGME practice-based learning and improvement requirements into reality: systems quality improvement projects as the key component to a comprehensive curriculum. *Quality and Safety in Health Care.* 2009; 18:217-224.
30. Gordan P, Kerwin TL. ACGME Outcomes Project: selling our expertise. *Family Medicine.* 2004; 36:164-167.

Annotated Bibliography

Lynch DC, Swing SR, Horowitz SD, Holt K, Messer JV. Assessing Practice-Based Learning and Improvement. *Teaching and Learning in Medicine.* 2004; 16(1):85-92.

This is a terrific general overview of assessment of PBLI. The authors promote the use of the portfolio as a collection of instruments that address the variety of learning content and thus can provide evidence of achievement in discrete areas and also provide a base for multi-faceted projects. The piece is a bit dated regarding the nature of some of the specific examples and suggested instruments, but the wide review of literature provides a valuable background for this subject.

Varkey P, Natt N, Lesnick T, Downing S, Yudkowsky R. Validity evidence for an OSCE to assess competency in Systems-based Practice and Practice-based Learning and Improvement: A preliminary investigation. *Academic Medicine* 2008; 83(8):775-780.

This well-designed pilot study of an OSCE to assess outcomes in the combined areas of PBLI and SBP is a must-read. The authors give specific guidance regarding tool development and discuss limitations involved with rater training and calibration of raters when using standardized stations to assess PBLI and SBP. Reviewing the variety of stations in the OSCE reinforces the concept of PBLI as a collection of discrete yet inter-dependent subcompetencies. The complexity of this OSCE may limit its widespread implementation as deployed in the small cohort reported. Don't let that deter you from implementing parts of this very well-designed assessment.

Leenstra JL, Beckman TJ, Reed DA, Mundell WC, Thomas KG, Krajicek BJ, Cha SS, Kolars, JC, McDonald FS. Validation of a method for assessing resident physicians' quality improvement proposals. *Journal of General Internal Medicine.* 2007; 22(9):1330-1334.

Resident knowledge of practice-based learning and improvement learning elements involved in quality improvement (QI) projects can be systematically assessed using a tool developed by Leenstra *et al*. This Quality Improvement Proposal Assessment Tool (QIPAT-7) has been used as a measure of quality improvement learning associated practice based learning elements submitted through a QI proposal and reported interrater reliability between 0.79 to 0.93 and internal consistency reliability among the items with

Cronbach's alpha = 0.87. Its basic design and practical approach to scoring makes this tool feasible and sensible.

10. Interpersonal and Communication Skills

Suzanne K. Woods, MD

"Listen to the patient. He is telling you the diagnosis."—Sir William Osler

The Competency Defined

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 health care team or other professional group.[1]

Rationale

Communication is a critically important part of practice that health care professionals need to master. Verbal and nonverbal communication occurs in all aspects of patient care and in working with colleagues and ancillary staff. In addition to the Accreditation Council for Graduate Medical Education (ACGME), several other international governing medical organizations have also recognized the critical importance of teaching and assessing interpersonal and communication skills (ICS) longitudinally during medical school and residency training.[2,3]

Good communication skills with patients are critical in the delivery of effective patient care.[2,4] Some of the earliest research on patient-physician communication was completed by Korsch and colleagues in the 1960's. Their observations of pediatric patient encounters allowed them to describe challenges with communication and to offer guidance to practioners on skill development in this fundamental competency.[4] More recently, since the original ACGME competencies were unveiled, an international expert consensus group of medical education leaders was convened to further define and expand the ICS competency. In 1999, 21 leaders convened in Kalamazoo, Michigan, to define a set of key elements in physician-patient communication. The conference was held to "identify and specifically articulate ways to facilitate communication teaching, assessment and evaluation."[5] Their work resulted in publication of "The Kalamazoo Consensus Statement" in 2001 which delineated the seven essential sets of communication tasks and will be discussed later in this chapter. The authors defined the general components of the ICS competency as follows:

- "Communication skills are the performance of specific tasks and behaviors such as obtaining a medical history, explaining a diagnosis and prognosis, giving therapeutic instructions, and counseling."[5]
- "Interpersonal skills are inherently relational and process oriented; they are the effect communication has on another person such as relieving anxiety or establishing a trusting relationship."[5]

It is helpful to start with these definitions and then explore how these skills relate to physician relationships. It has been shown that communication is a basic skill which can be taught and learned and is required at all levels of medical training.[2] Good communication and interpersonal skills facilitate patient symptom improvement, increased patient adherence to treatment plans, better management of chronic conditions, increased patient and provider satisfaction, and a reduction in medical errors and malpractice claims.[2] For both the delivery of quality health care and effective day-to-day working relationships in all settings, effective communication and interpersonal skills need to be developed and assessed during medical training.

Goals

1. Acknowledge the importance of the interpersonal and communication skills competency as it relates to pediatric residency education and appreciate the universality of these skills in caring for patients, and communicating with colleagues in different venues.
2. Understand the assessment of ICS is a necessary part of a training program and should be conducted longitudinally to aid individual residents in successfully achieving this competence.
3. Identify assessment instruments and strategies that can measure trainee interpersonal and communication skills.
4. Discuss the elements of ICS, highlighting how these elements impact the delivery of care.

Case Examples

Case 1

One of your interns is on an emergency department (ED) rotation. You receive the end of rotation evaluation from the ED attending. It notes unsatisfactory scores in the Interpersonal and Communication Skills competency. Comments include that the intern often did not provide accurate history and physical exam information in both the ED notes and verbal presentations. Specifically, the history was often superficial and reports of physical exam findings were incomplete or erroneous. In addition, critical lab reports were excluded from the documentation and presentation. The attending found it challenging to trust the intern's diagnosis, management and plan, given the inaccuracies.

Case 2

A senior resident sends you a "fast feedback" email to share what a great job a colleague did on call the night before. The junior resident receiving the praise was working in the emergency department (ED) and cared for a medically complicated child well known to the medical center. The resident completed a rapid and focused evaluation of the child and identified the key medical issues. She then communicated with the subspecialty providers and general pediatrician who routinely participate in the care of the child. Following that, the ED resident gave a clear sign out to the accepting ward team to ensure continuity of care. While the team was completing their evaluation of the patient upon arrival to the ward, the parents complimented the ED resident citing her attention to detail and caring attitude toward their child. They stated this was the most seamless encounter in the ED to date they have experienced. They were well prepared by the ED resident and knew what to expect during the hospitalization. The parents even knew the names of the physicians on the care team that they would meet on the ward!

As the program director, you need to address these issues with the respective trainees and contemplate how to do so.

Points for Consideration

<u>What are essential elements of interpersonal and communication skills and how do they impact the physician-patient relationship?</u>

Effective communication is a critical skill in providing good patient care. A solid and effective physician-patient relationship, identified by quality communication, can result in improvements overall in patient health outcomes and specifically in patients' satisfaction, treatment compliance, and quality of life.[2,6] Suboptimal communication skills however, have been associated with undesirable clinical consequences and "physician burnout, professional dissatisfaction and increased litigation".[6]

Two key statements that program directors should be familiar with that address ICS are the Kalamazoo report mentioned previously and the Macy Initiative in Health Care Communication.

The Kalamazoo Consensus Statement identifies seven sets of essential communication tasks which should be taught and assessed:[5]

1. Build the doctor-patient relationship
2. Open the discussion
3. Gather information
4. Understand the patients' perspective
5. Share information
6. Reach agreement on problems and plans
7. Provide closure

These skills are closely linked to successful relationships with patients. This report addressed how a physician's competence in communication with patients can be assessed. It also highlights how interpersonal skills (IS) "build on communication skills" and identifies the key elements of IS:[5]

1. Respect
2. Paying attention to the patient with open verbal, nonverbal, and intuitive communication channels
3. Being personally present in the moment with the patient, mindful of the importance of the relationship
4. Having a care intent

Another key perspective on ICS comes from the Macy Initiative in Health Care Communication, a collaborative effort between three institutions to develop a communications curriculum in undergraduate medical education with the goal of improving physicians' communication skills. This three year project, started in January 1999, identified three broad categories of core communication skills: communication with the patient, communication about the patient, and communication about medicine and science.[5,7] It demonstrated that communication skills can be taught in "an effective and meaningful manner".[5]

The interpersonal and communication skills competency emphasizes not only ICS skills in interactions with patients, but also with other members of the health care team. In pediatric residency training, one must be able to communicate with others during teaching sessions, while working in care teams, in the process of handoffs, when running a code, and when interacting across different settings such as clinic, the ED and the inpatient facility. In Case 1, there was a deficiency in the gathering and sharing of information, inability to generate a clear evaluation and care plan, and subsequent challenges in the supervisor/trainee relationship. This can lead to problems in the delivery of safe and quality patient care. In Case 2, the model resident facilitates communication in all aspects and highlights how successful communication should occur on many levels.

Can interpersonal and communication skills be improved with education and assessment?

Multiple models for communication have been developed and can be used for education and assessment of the ICS competency. While the focus of this primer is on assessment and not teaching, there is a dynamic relationship between teaching and assessment of ICS. Assessment can drive learning and general teaching strategies for both education and assessment of ICS include role playing, direct observation, workshops focused on communication, case discussions, and self-reflection. These tools can be used in a variety of residency program settings. Residency training program directors need to provide formal communication skills training. This can be in the context of communication with patients, patient caregivers, nurses, other ancillary staff, faculty, and peers. Residents need to be comfortable with communication skills because high self-efficacy is related to successful use of these skills.[2]

Most medical school patient encounters are with adult patients; consequently, baseline pediatric communication skills in trainees entering pediatric training programs may not be well-developed.[4] Many educators feel this deficiency needs to be addressed more rigorously as pediatric interactions are unique and can be very difficult due to the development and cognitive stages of a child, the need to involve family and other caregivers with consideration of family dynamics, the legal system, and other factors.[4] Rider and colleagues have identified the following areas as challenging to trainees and warranting additional attention: ability to discuss end-of-life issues with patients and family members, speaking to children about serious illness, delivery of bad news, interacting with difficult patients and parents, cultural awareness/sensitivity, understanding psychosocial aspects of patient care, and understanding patients' perspectives on their illness.[2] Additionally, sensitivity to cultural differences, language barriers, and health literacy should also be taught and assessed during training. Residents need to be evaluated on their abilities to collaborate, manage conflict, and compromise in the health care setting.

Work done by the 2003 Harvard Macy Institute Program for Physician Educators led to expansion of the ACGME ICS. These efforts resulted in a list of 20 subcompetencies for ICS, based on the original three described at the beginning of this chapter.[8,9] This working group also developed a teaching toolbox for ICS at all level of medical training including medical students, residents, and faculty and listed teaching strategies that can be used at each level. Subsequently, the Academic Pediatric Association created guidelines specifically for pediatric residency educators that addressed the ICS competency.[2]

It is very important for residency training programs to develop effective methods for assessment of communication skills during all phases of training. Sequencing and building upon skill sets that are appropriate and relevant to a trainee's level of experience are critical in helping learners achieve competence. For example, one must first master the skills to conduct a medical history before being able to discuss end-of-life issues with a patient. However, questions have arisen regarding how to translate the competencies into specific skills and clinical actions that can be taught and assessed.

Program directors can assess ICS in several ways which may lead to improvement in resident's skills. For example, ICS includes writing legible, complete, accurate, and timely medical records which can be assessed using chart review. Program directors can also assess effective handoffs and sign-out procedures observing both verbal and written communication. The intern in Case 1 would benefit from ongoing feedback and assessment. The use of multi-source feedback, self-assessment, direct observation using checklists for evaluation of history and physical exams skills would be a starting point. This assessment can then aid in the remediation of the learner. The resident in Case 2 would benefit from feedback as well, stressing what communication and interpersonal skills she has that led to a successful patient encounter. Self-reflection and assessment even in cases where the outcome is good can be helpful for continued improvement. This resident can also be challenged to teach other learners effective ICS one-on-one or in team encounters with colleagues. Longitudinal evaluation of both trainees is important to ensure that each

achieves and maintains competence in ICS. The definition of competence in ICS grows as residents progress through training. Initial abilities to perform generic communication tasks expand to encompass successful performance in complex, demanding, and specialty-specific situations. This is illustrated by the intern's elementary skills in Case 1 and the more advanced skills of the resident in Case 2.

What resources are available to assess residents in interpersonal and communication skills?

In 2002, the Kalamazoo group reconvened to review methods and tools for assessment of physician-patient communication.[13] The Kalamazoo II report, published in 2004, discusses the critical importance of assessment of the ICS competency, and describes and evaluates the use, cost, and evidence for specific assessment methods.[5] Tools and methods that can be used for assessment of the ICS competency include:

- Checklists
- Patient surveys
- Simulated patients
- Video/audiotapes
- Self-reflection
- Case discussions
- Empathy and emotional intelligence scales
- Role modeling/role play
- Multi-source evaluations
- OSCEs

Use of several different measures is recommended for assessment of the ICS competency. This will provide the most reliable and valid evidence of successful competency achievement. Some of the assessment tools mentioned above are described in greater detail in the paragraphs below.

Checklists: These are the most frequently used assessment tools in residency training programs. They allow an observer to rate the performance of a trainee in several communication behaviors. A numeric scale is used for rating and the checklist can have anchoring statements for each number to describe the behavior more specifically and delineate what satisfactory performance involves. Checklists, depending on who uses and completes them, can be both a form of formative and summative assessment.

Patient surveys: These are very important in the assessment of ICS, as the assessor is personally involved in the interaction and relationship with the physician. Unlike a third party observer, the patient can address many components of the physician's interpersonal and communication skills such as professionalism and humanism.[3] Patient assessments can complement faculty assessments of a learner. Of note, there is some data which correlates a patient's survey rating for a physician with the patient's perceived health status. Physicians tend to receive higher ratings from patients in good health than from patients in poor health.

Self-assessment: Self-assessment can be used in many areas of graduate medical education. For example, in the assessment of how a learner delivers bad news, the resident can be asked to complete a self-assessment of previous patient communication skills training and past experiences of giving bad news to patients and caregivers. Other activities, such as the use of standardized patients or participation in a workshop focused on the delivery of bad news and discussions of end-of-life care could then be used to improve areas of communication where the learner feels less confident or deficient.

Standardized/simulated patients (SP): Standardized patient experiences for evaluation of communication skills have been in use for many years. These patients may use checklists or rating scales to assess the learner. Patient assessment can be paired with either direct observation by a faculty member in real time, or review of the video-taped encounter, using a checklist to assess the encounter. Review with the learner can then follow this exercise to provide timely feedback. A key to the use of SP's is that these are patient-centered evaluations and assess communication skills well. Disadvantages include the cost of this tool and the need to have enough staff to train the standardized patients. Another use of simulated patients is in the objective standardized clinical examination (OSCE). In an OSCE, SP's play a specific role and use checklists to rate the trainee. Scenarios may focus on a variety of skills, such as the delivery of bad news, specific clinical examination skills, counseling patients about risk factor modification or assessing pain management in cancer patients.

Multi-source (360 degree) Evaluations: This assessment tool involves many individuals from a full circle (360 degrees) who evaluate a learner. This can assessments by include faculty, nurses, peers, ancillary staff including pharmacists and social workers, medical students, and clerical staff, in addition to a self-assessment. It is important to collect this feedback in a timely manner following interactions the above individuals have with a specific learner. This feedback should be shared with the trainee in aggregate and anonymous fashion by the program director, ideally during semiannual meetings. The advantage of this approach is that many evaluators participate, which helps to increase data validity and reliability. Also, ideally the learner's self-assessment skills will improve with the availability of outside assessments. Disadvantages include the need for many evaluators and the ability to manage the data that is collected.

There are many tools to aid a residency training program director in the assessment of interpersonal and communication skills. It is also essential to assess a trainee when individual concerns are raised with regards to interpersonal skills and deficiencies in communication and the methods above can assist with that assessment.[3] There are other techniques that can be employed to assess components of ICS such as medical interviewing, delivering bad news, understanding of issues of disparities, medical errors, and access to care. These are described in several resources including the ACGME Outcomes Project.[1] Other specific tools that can be considered for assessment are outlined in the table on the following page.

Another concept gaining popularity that may aid in the understanding and assessment of the ICS competency is *Emotional Intelligence* (EI). EI is broadly defined as one's self-management and interpersonal skills.[7] EI is a set of four distinct but related abilities that include perceiving emotions, using emotions, understanding emotions, and managing emotions. Some characteristics of emotional intelligence include stress tolerance, adaptability, *empathy*, impulse control, optimism, and problem-solving skills.

It has been suggested that the theory of EI may help to define the "specific abilities and complex processes" of the ICS competency.[7] Subsequently, this can lead to a better understanding of how to teach and assess ICS in residency training. Measurement of EI includes two domains: self-report and ability measures. Available tools for self-report from outside of the medical field include the Bar-On Emotional Quotient Inventory (EQi) and the Self-Report Emotional Intelligence Test (SREIT). The concern with such self-assessment tools however, is that they "reflect perceptions of emotional abilities rather than measures of the abilities themselves".[7] Therefore tests of ability measures need to be paired with self-assessment. An example of a tool is the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT). This test specifically uses tasks to assess the 4 components of EI. Further work is needed to determine how the four abilities of EI relate to interpersonal and communication skills in graduate medical education.

Assessment Tool	Description	Strengths	Weaknesses
CCOG Calgary Cambridge Observation Guides	Initially designed to facilitate teaching in communication skills, then adapted for assessment Case specific content checklist is paired with the tool1	Measures process and content of the medical interview Five point plan allows for structure to individual communication skills	Difficult to use Lengthy
Wy-Mii Wayne State Medical Interviewing Inventory	Designed to assess residents' ICS in the context of a real or standardized ambulatory medical interview[10] 27 item instrument	Aims to differentiate between high and low levels of interviewing skill level Faculty can focus on a smaller number of teaching points for each residency year	Not clear how well tool can discriminate across the full range of skill levels
SEGUE Framework **S**et the Stage **E**licit Information **G**ive Information **U**nderstand the patient's perspective **E**nd the encounter	Checklist used by faculty to evaluate medical interviewing skills Originally designed to evaluate medical students Focuses on specific, objective communication "tasks" and assesses if a learner has completed each task[10,11]	In use for many years Internal consistency Inter-rater reliability Easy to use	Checklist limits measurement of the quality of the interview Does not measure context of applied task Faculty must be trained to use the evaluation tool Highly content specific[11]

Three existing tools for assessing interpersonal and communication skills in medicine.

Competence in interpersonal and communication skills is critical in the development of a physician. These skills must be "learned, practiced and maintained, like all other essential skills of the clinician".[16] Effective communication can result in increased patient and physician satisfaction, improved quality of care, increased patient compliance with treatment plans, reductions in medical errors, fewer lawsuits, and better management of chronic diseases.[2,6,8,16] It is important to assess these skills longitudinally during residency training to ensure competence.

Lessons Learned

- Communication is important in virtually all aspects of patient care.[5]
- Communication skills in pediatrics are different, unique and more complex when compared with adult medicine and necessitate different training of learners.
- Program directors need to teach and assess individual residents' confidence with these skills.
- The definition of competence in ICS grows as residents progress through training. Initial abilities to perform generic communication tasks expand to encompass successful performance in complex, demanding and specialty-specific situations.[3]
- Methods need to be developed to allow for better training and promotion of effective communication skills for residents. Paired with this education must be assessment, feedback, the opportunity

for trainee self-assessment, and continued promotion and development of this competency over time.

References

1. Accreditation Council for Graduate Medical Education. The ACGME Outcome Project. *Advancing Education in Interpersonal and Communication Skills*. 2005. [Internet] *http://www.acgme.org/outcome/implement/interpercomskills.pdf*. Accessed May 18, 2011
2. Rider EA, Volkman K, Hafler JP. Pediatric residents' perceptions of communication competencies: Implications for teaching. *Medical Teacher* 2008; 30:e208-e217.
3. Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, All Participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing Competence in Communication and Interpersonal Skills: The Kalamazoo II Report. *Academic Medicine* 2004; 79(6):495-507.
4. Dube CE, LaMonica A, Boyle W, Fuller B, Burkholder GJ. Self -assessment of communication skills preparedness: Adult versus Pediatrics Skills. *Ambulatory Pediatrics* 2003; 3(3):137-141.
5. Participants in the Bayer-Fetzer Conference on Physician-Patient Communication in Medical Education. Essential Elements of Communication in Medical Encounters: The Kalamazoo Consensus Statement. *Academic Medicine* 2001; 76(4):390-393.
6. Farber NJ, Urban SY, Collier VU, Weiner J, Polite R, Davis E, Boyer EG. The good news about giving bad news to patients. *Journal of General Internal Medicine* 2002; 12:914-922
7. Grewal D, Davidson H. Emotional Intelligence and Graduate Medical Education. *JAMA* 2008; 300(10):1200-1202.
8. Kalet A, Pugnaire MP, Cole-Kelly K, Janicik R, Ferrara E, Schwartz M, Lipkin M, Lazare A. Teaching Communication in Clinical Clerkships: Models from the Macy Initiative in Health Communications. *Academic Medicine* 2004; 79(6):511-520.
9. Rider EA, Keefer CH. Communicating skills competencies: definitions and a teaching toolbox. *Medical Education* 2006; 40:624-629.
10. Skillings JL, Porcerelli JH, Markova T. Contextualizing SEGUE: Evaluating Residents' Communication Skills within the Framework of a Structured Medical Interview. *Journal of Graduate Medical Education* 2010 Mar; 102-107.
11. Makoul G. The SEGUE Framework for teaching and assessing communication skills. *Patient Education and Counseling* 2001; 45:23-34.
12. Buckman RA. Breaking Bad news: the S-P-I-K-E-S strategy. *Community Oncology* 2005 Mar/Apr:138-142.
13. Lin C-T, Platt FW, Hardee JT., Boyle D, Leslie B, Dwinnell B. The Medical Inquiry: Invite, Listen, summarize. *Journal of Clinical Outcomes Management* 2005; 12(8):415-418.
14. Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, Till J. Doctor-patient communication: the Toronto consensus statement. *British Medical Journal* 1991; 303:1385-1387.
15. Donnelly MB, Sloan D, Plymale M, Schwartz R. Assessment of Residents' Interpersonal Skills by Faculty Proctors and Standardized patients: A Psychometric Analysis. *Academic Medicine* 2000; 75(10):S93-S95.
16. Gordon G. Defining the Skills Underlying Communication Competence. *Seminars in Medical Practice* 2002; 5(3):21-28.

Annotated Bibliography

Accreditation Council for Graduate Medical Education. The ACGME Outcome Project. Advancing Education in Interpersonal and Communication Skills. June 5, 2005. [Internet] *http://www.acgme.org/outcome/implement/interpercomskills.pdf*

The ACGME website has many educational resources available to program directors. Specific to the ICS competency, this resource offers a description of ICS, and has sections that address teaching and assessment of ICS, specific tools and an FAQ section. It is loaded with additional references including books, journal articles and web based guide links. All program directors should be familiar with this reference for the competencies.

Rider EA, Volkman K, Hafler JP. Pediatric residents' perceptions of communication competencies: Implications for teaching. *Medical Teacher* **2008; 30:e208-e217.**

Pediatric communication competencies are unique yet little is known about pediatric residents' perceptions regarding these skills. The authors developed a 47-item cross-sectional questionnaire to study residents' attitudes and perceptions regarding ICS. The residents felt competency in communication was important, especially being able to effectively communicate with their patients and demonstrating empathy and caring. However, they reported a lack of confidence in advanced communication skills specifically ability to discuss end-of-life issues, speaking with children about serious illness, dealing with difficult patients, and cultural awareness/sensitivity. The authors recommend longitudinal training in core and advanced communication skills.

Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, All Participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing Competence in Communication and Interpersonal Skills: The Kalamazoo II Report. *Academic Medicine* **2004; 79(6):495-507.**

The initial Kalamazoo Consensus Statement published in 2001 identified seven essential elements of communication tasks. These general elements provided a useful framework for communication-oriented curricula and standards. The follow up work by this group of educators summarized the state of the art in teaching and assessing competence in ICS. Specifically this report reviews three methods for assessment of ICS: (1) checklists of observed behaviors in interactions, (2) surveys of patients' experience in interactions, and (3) examinations using oral, essay, or multiple-choice response questions. Their table on page 501 highlights key features of five ways to assess ICS including: the assessment strategy; use in settings such as UME, GME, CME; costs, comments on the strategy and a reference list for each one. Great overall review of the assessment of ICS.

Kalet A, Pugnaire MP, Cole-Kelly K, Janicik R, Ferrara E, Schwartz M, Lipkin M, Lazare A. Teaching Communication in Clinical Clerkships: Models from the Macy Initiative in Health Communications. *Academic Medicine* **2004; 79(6):511-520.**

Medical educators are faced with the challenge of how to instruct learners in effective communication. To address this issue, New York University School of Medicine, Case Western Reserve University School of Medicine and the University of Massachusetts Medical School collaborated to develop, implement, and evaluate a comprehensive communications skills curriculum. The project, funded by the Josiah P. Macy, Jr. Foundation, took the name "the Macy Initiative in Health Communication". A variety of teaching methods were used for third year medical students. The article describes how each school tailored the cur-

riculum to fit its individual needs. This work may help program directors and clerkship directors develop cross-disciplinary skills curricula.

Gordon G. Defining the Skills Underlying Communication Competence. *Seminars in Medical Practice* **2002; 5(3):21-28.**

The author does a great job of defining the skills underlying the ACGME ICS competency. He provides an overview of many educational models that program directors can then research in more depth. He explains the skills related to the ICS competency such as: (1) create and sustain a therapeutic relationship, (2) demonstrate caring and respectful behaviors, (3) use effective listening skills, (4) gather essential and accurate information, and others, with reference tables that address each area succinctly. He addresses challenges to trainees as well as educators. A very well-rounded examination of this competency and recommended reading for all educators.

11. Professionalism

Stephen Ludwig, MD

"Professionalism is knowing how to do it, when to do it, and doing it."   - Frank Tyger

"A professional is someone who can do his best work when he doesn't feel like it." - Alistair Cooke

The Competency Defined

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, altruism, honesty, compassion and integrity.
- Demonstrate a commitment to ethical principles.
- Demonstrate sensitivity and responsiveness to patients' culture, age, gender and disabilities. [1]

Rationale

Professionalism is an important and often overlooked area in resident assessment. This competency is one that often only comes to the attention of a program director when there is a lapse in professionalism or a series of professional difficulties lead to outcomes that reveal a pattern.[2, 3] Unfortunately, when glaring professionalism infractions, such as drug use, financial impropriety, or even homicide, come to the attention of the public, the medical profession as a whole suffers a loss of trust. Often it is only after such a scandal that hindsight reveals more minor professionalism lapses that were never corrected. Like other core competencies, professionalism should be assessed on a regular basis in every resident. This chapter explores some of the assessment methods available to program directors.

Some question whether professionalism can be taught; they hold that the traits and behaviors that constitute professionalism are learned (or not) in the first six years of a child's life. However, although parents may instill early concepts of proper behavior and ethical conduct, program directors and faculty translate these traits into the context of the medical setting and encourage the formation of habits that are important to behaving as a professional pediatrician.

The American Board of Pediatrics (ABP) and the Association of Pediatric Program Directors regard professionalism as an important topic and have issued a guide entitled *Teaching and Assessing Professionalism*.[4] This excellent resource can be found on the American Board of Pediatrics website at *http://www.abp.org*. The guide offers program directors a format for developing educational sessions to engage trainees in discussion and self-reflection exercises about professionalism. The guide attempts to help program directors answer three questions: (1) What are the important elements of professionalism? (2) How can expectations regarding professional conduct be communicated to pediatric trainees? (3) What methods are appropriate for assessing professionalism during residency training?

Professionalism is not a have or have not competency like a checklistable procedural skill. According to David T. Stern, a scholar of medical professionalism, "Professionalism is not what you do every time, but what you do over time."[5] This suggests the importance of creating habits of behavior that will be life-long and endure even in stressful situations. There is a developmental sequence of professional behaviors that

should mature over time and experience. The expression of professionalism may also vary by context such as expectations based on level of responsibility and duty, coincident stressors, and other situational factors. Professional behaviors also change with broader changes in society; for example, the emergence of electronic networking and social media have created new professionalism challenges.

Goals

1. Understand the breadth of the concept "professionalism" as it relates to assessment.
2. Appreciate that routine assessments of professionalism must take place even absent the occurrence of serious professionalism lapses.
3. Be aware of the tools that can be employed to assess this competency.
4. Understand the process of assessing professionalism and your obligation as a program director to assess professionalism in each resident and report the results to the ABP.

Case Examples

Case 1

An intern making family-centered rounds introduces himself to the anxious and upset parents of a patient admitted to rule out leukemia, "Hi, I'm Jim. I'll be working with Dr. Jones."

Case 2

A nurse from the Emergency Department calls you. She reports a confrontation between a pediatric resident and an orthopedic resident that occurred in the Emergency Department in the middle of the night.

Case 3

Residents in your program vote on an annual award for the resident who demonstrates humanitarian behaviors and outlook. Twenty out of twenty-five residents vote for Bill, a PGY2.

Case 4

On your global evaluation form you are asked, "How do you evaluate this resident's cultural competence: unacceptable, fair, good, excellent, or outstanding?" You have spent one week on-service with this resident and wonder how to respond.

Points for Consideration

What is professionalism? What are its components?

There are at least four components of professionalism, each of which should be assessed:

- *Professionalization* is taking on the mantle of a physician. It is a transition from medical school—where one is a student of medicine—to residency—where one is a doctor and responsible for human life and health. It is the assumption of a professional role.
- *Professional conduct* is the way we act with patients, parents, hospital staff, and colleagues. These are the behaviors expressed in day-to-day interactions.
- *Humanism* is a set of behaviors that convey caring, empathy, sense of duty, compassion, and altruism. Humanism stresses not the distinction of being a professional but the sameness of being

another member of the human family. In the words of AAMC President Emeritus Jordan Cohen, "humanism provides the passion that animates authentic professionalism."[6]

- *Cultural competence* is the ability to understand people from other backgrounds, places, or cultures than the resident's own in order to interact in ways that are appropriate and consistent with the values and traditions of others.

How do you assess professionalization?

In the transition from student to graduate physician, a trainee develops her professional identity. Through direct observation or multi-source feedback one can gauge her stage in this process. One encounter should never lead to a final determination. However, in Case 1, the attending should be alerted to the fact that this intern introduces himself by first name only. He may not yet understand the context of this encounter and the need for these parents and their child to know they are cared for by professional staff, not just a friendly "Jim" or "Jill". As an intern, this trainee may not yet see himself as a doctor. It might be helpful to know how others see him: as a doctor or just a young man helping a doctor. Parent and peer evaluations may add to the attending physician's formulation of the intern's professional development.

What factors contribute to lapses in professional behavior?

The report by the nurse in Case 2 is a *critical incident*. It should be documented and saved in the resident's portfolio. Perhaps it will be the only one of its kind; perhaps it will be one of many. Critical incidents both positive and negative paint a picture of professional conduct. Incidents may include intra-professional conflicts, interactions with families, or relevant behaviors outside the workplace. A trainee's misbehavior outside the hospital may reflect on his professional status particularly if it comes to public attention. In further assessing this resident, the program director should dissect the incident and identify factors at play. Why the conflict? What other means of resolution were tried? Was fatigue, anxiety, depression, or another stressor present? Critical incident reporting is only the starting point to understanding and assessing professional behavior.

How does one assess humanism?

Daily interactions with a resident give insight into the resident's humanism. One can assess the trainee on the components of humanism or use a more global evaluation. Awards, as in Case 3, are a form of peer and/or faculty recognition of key traits. In this case, the resident stands out as exceptional. But all residents must be evaluated on their humanistic traits on a regular basis. Self-reflection, observations of home visits, and questions at rounds and in conferences are all ways to assess humanism. There are also standardized measures of some humanism traits such as the Jefferson Empathy Scale.[7]

How can we assess cultural competence?

Cultural competence is the fourth element of professionalism. Like the others, it cannot be judged with a Likert scale item or in a single situation. Global assessments of cultural competence can be improved by providing raters with graded anchors arranged along a developmental sequence, such as:

1. Seems unaware of cultural differences.
2. Makes judgments based on his/her own background and experience.
3. Identifies cultural differences and how they might impact clinical care.
4. Develops educational materials to train one's self and others about important cross cultural issues.

Case 4 highlights the question of whether an attending can assess a resident's proficiency in cultural competence during a one-week rotation. The answer is clearly "no". All aspects of professionalism are best judged over time in a longitudinal preceptor relationship.

What are the tools that can be used to evaluate professionalism and what are the strengths and weakness of each?

Several methods and tools have been used to assess professionalism:[8,9]

Critical incidents: These are reports that are filed by hospital personnel, family members, and others. They may come to the resident directly or to the program director. They may be either positive or negative. They should be reviewed by the program director, discussed with the resident, and maintained in the resident's portfolio. Hickson and colleagues have described the relationship between patient complaints and malpractice risk.[10]

Peer assessment: These structured assessments should ask questions related to each of the four areas of professionalism. They are completed anonymously, analyzed by the program director, and used to give formative feedback to residents. These are often highly reliable as peers have a keen sense of whom they want to work with and with whom they would prefer not to be teamed.

Professionalism mini-clinical evaluation exercise: A mini-CEX focused on professionalism may be conducted with real patients or simulated patients.[11] It is critical to have a standardized rating form and faculty development for raters so that the scores have sufficient inter-rater reliability. Professionalism may also be assessed in the context of assessments of other skills and competencies, as it is an expected component of all practice.

Multi-source assessment: A 360 degree evaluation asks attending staff, nursing, peers, and patients to each evaluate a resident's professionalism.[12,13] However, a recent review of assessment methods for ACGME competencies found little benefit to multi-source assessment.[14] Each rater may have her own view, so understandable common anchors that describe behaviors should be used to help standardize ratings.

The feasibility, reliability, and validity of each of these methods are outlined in the table on the following page.

Other assessment modalities may also be helpful. Veloski, et al[15] reviewed 114 instruments reported in the literature that might be used to measure professionalism. They concluded that although there were many published tools, few had documented evidence of reliability, validity, and feasibility. Recently, the National Board of Medical Examiners[16] has launched an effort around assessment of professionalism. They have developed a multisource feedback tool that measures professional behaviors. See *http://www.nbme.org/apb* for details and materials.

What evidence is there that assessing professionalism can make a difference? Are there risk factors that can predict later professionalism problems?

Papadakis and colleagues[16] have studied whether performance measures made during residency predict the likelihood of future disciplinary action against internists. The Residents' Annual Survey rating and the American Board of Internal Medicine (ABIM) examination scores were used as predicators. Both of these measures were shown to be predictors of later problems that required disciplinary action by state licensing boards. In another study performed by Hodgson and colleagues[18] the relationship between measures of unprofessional behaviors of medical students and later disciplinary action were explored. The California Psychological Inventory (CPI) had been used as a predictor for law enforcement officers and in this study

students with scores indicative of irresponsibility, lack of self-improvement, and poor initiative tended to have unprofessional behavior.

What can be done about unprofessional behavior?

There are many efforts being made for remediation of unprofessional behavior. Many program directors have struggled to find ways to improve resident behaviors. Hauer and colleagues[19] review the existing literature on this topic. There were thirteen published studies which were mainly small, descriptive, and single institution based. The authors of these studies suggest a multi-pronged model that includes frequently used multiple assessment tools, individualized instruction, deliberate practice followed by feedback, reflection, and reassessment. The authors call for more work in this area. Remediation is the goal of professional assessments that show deficiencies. Having more effective remediation strategies will drive more routine use of assessment tools.

Assessment Method	Feasibility	Reliability and validity	Notes
Critical Incidents	Low-cost evaluation; can be done on paper or via a Web-based system; faculty development would facilitate use	Studies document correlation with discipline by state medical board for students about whom serious concerns were raised; no good data on reliability	Consider use of Praise/Early Concern Card developed by the ABIM
Peer Assessments	Resources required for distributing and collecting data; will add to overall evaluation response burden for residents	Six to eleven evaluations can produce reliability coefficient of 0.7; multiple sources of validity evidence	Involving residents in developing the instrument can increase buy-in
Clinical Evaluation Exercises	Relatively easy to implement after initial training with the form; requires faculty time for observation of resident-patient interactions	Use of ten to twelve raters provides a reliability coefficient of 0.8; good evidence of validity;	Covers the full range of professional behaviors
Multi-source Assessments	High-cost evaluation; requires system to distribute and collect from a variety of sources; need to develop database to analyze data	Requires a large number of patient evaluations for high-stakes decisions, but well-suited to formative feedback	NBME instrument is a promising tool, but results not yet available

Methods for assessing professionalism. Adapted from The American Board of Pediatrics and The Association of Pediatric Program Directors. *Teaching and assessing professionalism: a program director's guide*. American Board of Pediatrics, 2008

Lessons Learned

- Professionalism is an important competency often taken for granted unless there is an untoward report.
- Program Directors may wish to assess individual elements of professionalism (e.g. cultural competence, humanism) rather than attempting to globally assess professionalism as a single construct.
- Several methods and tools for assessing professionalism exist but each has both strengths and weaknesses.

- It is important to make active, prospective assessments of professionalism; do not wait to find professionalism lapses at the end of the training cycle.
- In assessing professionalism one must examine the assessments of many individuals and the context in which these assessments are made.

References

1. Accreditation Council for Graduate Medical Education. Outcome project—Practical implementation of the competencies. 2006. [Internet] *http://www.acgme.org/outcome/*. Accessed May 18, 2011.
2. Papadakis MA, Loeser H, Healy K. Early detection and evaluation of professionalism deficiencies in medical students: One school's approach. *Academic Medicine* 2001; 76:1100-1106.
3. Dyrbye LN, Massie Jr FS, Eacker A, Harper W, Power D, Durning SJ, Thomas MR, Moutier C, Satele D, Sloan J, Shanafelt TD. Relationship between burnout and professional conduct and attitudes among US medical students. *JAMA* 2010; 304:1173-1180.
4. The American Board of Pediatrics and The Association of Pediatric Program Directors. *Teaching and assessing professionalism: a program director's guide*. American Board of Pediatrics; 2008.
5. Stern DT, editor. *Measuring Medical Professionalism*. New York, NY: Oxford University Press; 2006.
6. Cohen JJ. Linking professionalism to humanism: What it means, why it matters. *Academic Medicine* 2007; 82:1029-1032.
7. Hojat M, Gonnella JS, Mangione S, Nasca TJ, Magee M. Physician empathy in medical education and practice: Experience with the Jefferson Scale of Physician Empathy. *Seminars in Integrative Medicine* 2003; 1:25-41.
8. Whitcomb ME. Fostering and evaluating professionalism in medical education. *Academic Medicine* 2003; 77:473-474.
9. Lynch D, Surdyk P, Eiser A. Assessing professionalism: A review of the literature. *Medical Teacher* 2004; 26:366-373.
10. Hickson GB, Federspiel CF, Pichert JW, Miller CS, Gauld-Jaeger J, Bost P. Patient complaints and malpractice risk. *JAMA* 2002; 287:2951-2957.
11. Cruess RL, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. The professionalism mini-evaluation exercise: A preliminary investigation. *Academic Medicine* 2006; 81:S74-S78.
12. Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, DeWitt TG, Britto MT. Effect of multisource feedback on resident communication skills and professionalism. *Archives of Pediatric & Adolescent Medicine* 2007; 161: 44-49.
13. Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. *American Journal of Physical Medicine & Rehabilitation* 2003; 82:394-402.
14. Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the Accreditation Council for Graduate Medical Education: A systematic review. *Academic Medicine* 2009; 84(3):301-309.
15. Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: A review of studies with instruments reported in the literature between 1982 and 2002. *Academic Medicine* 2005; 80:366-370.
16. Assessment of professional behaviors—A new service of the NBME for residency and fellowship programs. [Internet] The National Board of Medical Examiners. Webinar, June 7, 2010. *http://www.nbme.org/schools/apb/PDF/APB_672010_webinarslides.pdf*. Accessed May 18, 2011.

17. Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Annals of Internal Medicine* 2008; 148:869-876.
18. Hodgson CS, Teherani A, Gough HG, Bradley P, Papadakis MA. The relationship between measures of unprofessional behavior during medical school and indices on the California psychological inventory. *Academic Medicine* 2007; 82:S4-S7.
19. Hauer KE, Ciccone A, Henzel TR, Katsufrakis P, Miller SH, Norcross WA, Papadakis MA, Irby DM. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Academic Medicine* 2009; 84:1822-1832.

Annotated Bibliography

Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzalez del Rey JA, DeWitt TG, Britto MT. Effect of multisource feedback on resident communication skills and professionalism. *Archives of Pediatric and Adolescent Medicine* 2007; 161:44-49.

In this study residents were assigned to a multisource feedback group and completed a self-assessment form, received feedback from parents and nurses, and participated in a tailored coaching session as well as getting standard feedback. The study showed when compared to a control group of those who just received standard feedback the experimental group performed better in communication skills and professional behavior.

Cruess R, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. The professionalism mini-evaluation exercise: A preliminary investigation. *Academic Medicine* 2006; 81S:S74-78.

The authors created and studied the equivalent of a mini-CEX for professionalism called the P-MEX for Professionalism Mini-Evaluation Exercise. After a literature review, 142 observable behaviors were identified that embody professionalism. These were distilled into 24 and converted into an evaluation instrument similar to the mini-CEX. A 4 point scale was used where 1=not acceptable, 2=below expectations, 3=met expectations and 4=exceeded expectations. A fifth category was "not observed" or not "applicable." The intent was to use this in any setting in which behaviors could be directly observed. In addition to the 24 items there was space for comments and critical incidents. There were two copies so that one could be given to the student with feedback. A total of 211 forms were collected on 74 medical students. Item analysis, factor analysis, and a generalizability analysis were completed. The results showed that 3 questions were redundant, 3 factored into more than one category and needed to be reworded, and somewhere between 10 and 12 forms needed to be completed to achieve a dependability coefficient of 0.80. The four categories resulting from the factor analysis are: (1) doctor-patient relationships, (2) reflective skills, (3) time management, and (4) interpersonal and communication skills. Three of the four items that showed 3% or more of the ratings as "below expectations" fell under the category of self-assessment and these items were "demonstrated awareness of limitations", "solicited feedback", and "addressed gaps in own knowledge and skills."

Hojat M, Gonnella JS, Mangione S, Nasca TJ, Magee M. Physician empathy in medical education and practice: Experience with the Jefferson Scale of Physician Empathy. *Seminars in Integrative Medicine* 2003; 1:25-41.

The Jefferson Scale of Physician Empathy which consists of 20 Likert-type items was developed and tested in 1007 physicians. Three meaningful factors emerged: (1) perspective taking, (2) compassionate care, and (3) standing in the patient's shoes. The scores were seen to be internally consistent and relatively stable over time. The authors conclude that empathy is a multi-dimensional concept that varies among physicians but can be measured with a psychometrically sound tool.

Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: A review of studies with instruments reported in the literature between 1982 and 2002. *Academic Medicine* **2005; 80:366-370.**

This paper reviewed studies published from over a twenty year interval and included 114 papers which looked at specific elements of professionalism. Few studies addressed professionalism as a competence construct. The report concluded that there are few well documented studies of these instruments and that their use is formative and summative evaluation is limited. Authors advise that when looking at tools for this purpose the user must look critically at content validity, reliability, and practicality.

12. Systems-based Practice

Julia McMillan, MD

"The systems we work in often can be difficult to identify and define. Although we work in numerous systems all day, every day, it's difficult to 'see' a system. It's like asking fish to describe water—it's easier to be aware of the system when the system fails."[1]

The Competency Defined

Systems-based practice requires residents/fellows to demonstrate an awareness of and responsiveness to the larger context and system of health care, as well as the ability to call effectively on other resources in the system to provide optimal health care. Residents/fellows are expected to

- Work effectively in various health care delivery settings and systems relevant to their clinical specialty
- Coordinate patient care within the health system relevant to their clinical specialty
- Incorporate considerations of cost awareness and risk benefit analysis in patient care
- Advocate for quality patient care and optimal patient care systems
- Work in interprofessional teams to enhance patient safety and improve patient care quality; and
- Participate in identifying system errors and in implementing potential systems solutions.[2]

Rationale

Appropriate medical care involves moving beyond the dyadic relationship between physician and patient to an understanding that patient health depends upon a complex system made up of other individuals, institutional policies, and regulations. Competence in systems-based practice (SBP) incorporates both knowledge and performance. Effective use of systems to enhance care requires knowledge of the components of health care systems at all levels—within hospitals and clinics, in the community, and in the home. Knowledge of costs and of relative risks and benefits of management options is also essential. Moreover, the SBP competency requires that knowledge be used to continually enhance care, both by understanding how to work effectively within the systems that exist, and by working to improve those systems when necessary. "SBP is care that is sensitive to the context in which it is delivered."[1]

Methods for assessment of SBP are limited, in part because faculty members and medical educators are themselves uncertain about their knowledge of medical care systems, and in part because assessment requires standards for observed behaviors of trainees in the process of delivering care. These behaviors are not necessarily quantifiable and require us as educators to think about qualitative means of assessment. Although efforts of residents and fellows to utilize their knowledge of systems of care to improve the health of their patients are recognized on a daily basis by observant faculty and medical staff, only the patients and the families, who are the recipients of those efforts, may know their outcome.

Goals

1. Understand that the interplay between knowledge of the health system and the ability to work effectively within that system is essential for improving health.
2. Appreciate the importance of educating faculty evaluators about the components of SBP in preparation for their assessment of residents.

3. Recognize that multiple observers, including faculty, medical staff, peers, family, and residents themselves are needed for assessment of SBP.
4. Understand the importance of assessing interaction between the developing pediatrician and the complex system of health care, which can only be improved through the engagement of individuals.

Case example

The continuity clinic site for your residents is an inner city, hospital-based clinic, and many of the patients are children with special health care needs. The clinic staff has embraced the medical home as the model that will best meet the very complex needs of the patients being served. An important goal for your residents' continuity clinic experiences is that they develop an understanding of, and skill in, providing a medical home for their patients.[3] Components of this goal that you feel are important include the resident's knowledge of clinic operations, of the insurance plans available to the patients in the clinic, of mechanisms for referral of patients for additional care outside the clinic, and of the communities in which their patients live. You also want them to develop skill in coordinating the care of their patients, both within your clinic and when patients' health care needs require additional resources. In order to accomplish this goal it is critical that they learn to work effectively with the team of providers, including the nurses, clerical staff, social workers, and case managers. You recognize that faculty preceptors work closely with residents in overseeing their care of patients and that they are in a position to assess resident progress toward these goals, but you would also like to create additional methods for tracking residents' progress as they develop continuing relationships with their patients and coordinate the many aspects of their care. You would also like to assess individual resident engagement in efforts to identify and address deficiencies in the system of care.

Points for consideration

<u>How will you know that residents have learned what they need to know about the system of care in which they work?</u>

Much of the knowledge needed to provide coordinated care in a medical home is learned through seeking out the answers to questions needed in providing that care. However, information shared in a systematic manner through didactic sessions or on-line modules can provide basic knowledge that will be enhanced through active care of patients. Information provided by social workers, nurses, referral coordinators, lawyers, and home care agencies could be included in material presented to residents as a foundation for their responsibilities in continuity clinic.

Resident understanding of the roles of different health professionals as well as the systems in which care is provided can be assessed using both specific targeted testing and through subsequent direct observation. In one study that used web-based educational materials to teach and assess the competency of systems-based practice, modules about patient safety, error prevention, and systems theory, along with modules about the structure of the U.S. health care system were administered to medical students and residents.[4] Answers to pretest, mid-test, and posttest questions demonstrated both enhanced knowledge of the topics covered and retention of the information when tested one week following completion of the final module. This study was not conducted within a specific patient care setting, so the information conveyed was not likely to be immediately relevant to the students and residents, as would targeted information relevant to a particular continuity clinic setting in which residents were regularly caring for patients.

Direct observation by faculty preceptors and other health care providers (e.g., nurses and social workers) who work in the clinic setting can serve as an ongoing measure of resident knowledge about the system of

care. Assessment tools that incorporate descriptions of desired behaviors along a developmental continuum (anchors), rather than vague qualitative terms such as "good," "excellent," "superior," can help guide faculty members and other assessors. Some specific examples are given below.

Knowledge of health care systems					
Cannot Evaluate	Minimal apparent knowledge of current health care system for children	Needs to expand knowledge of systems for referrals, authorizations and role of primary care physician.	Understands impact of health care systems on individual patient care	Well-developed understanding of health care systems (public, private) as they relate to care of patients. Able to help patients navigate this system.	Detailed and sophisticated understanding. Able to partner with other providers in coordination of patient care across systems and along a continuum.
0	1	2	3	4	5
Patient advocacy					
Cannot Evaluate	No effort to aid patients in navigating complexities of health care system	Inadequate effort to investigate options and programs for patients	Accepts responsibility to seek resources for patient and arrange necessary follow-up	Able to do a thorough needs assessment for patients. Knows many available resources.	Unusually adept at seeking out help for patients and navigating systems for benefit of children and families.
0	1	2	3	4	5
Effectiveness as a primary care provider in the context of a medical home					
Cannot evaluate	Appears disinterested in coordinating care; plays little role in linking families to subspecialty and community resources	Exerts little effort in coordinating care for patients	- Develops supportive relationships with patients and families; is recognized by families as their primary care provider; identifies community and subspecialty resources for families	Consistent and effective provider of care for patients; advocates for patients beyond clinic visits; coordinates and facilitates services for families	Is consistently regarded by patients/families as their preferred provider; takes proactive role in coordinating care and linking patients to subspecialty and community resources; is cognizant of family dynamics, community support services and their impact on health
0	1	2	3	4	5

Examples of questions that could be used to assess the systems-based practice competency

How can residents demonstrate their application of knowledge of the health care system to their care for patients?

Assessment of this component of SBP requires input from multiple individuals who work with and observe residents in their care of patients. Again, assessment tools that target and describe specific behaviors along a developmental continuum can be used by faculty preceptors, nurses, social workers, and others who are in a position to observe each resident's prescribing patterns (to assess cost awareness and risk benefit analysis), appropriateness of referral to subspecialists and other resources, and level of communication with families.

As a part of the evaluation of resident competence in communication skills and professionalism, many residency programs have incorporated a survey of families. Most studies have found that parent assessment of residents is so uniformly positive that their feedback is useful only to reinforce positive behavior.[5,6] One study included questions about medical management in the parent questionnaire and compared parental responses to those of preceptors and of residents themselves.[6] Though parent assessment in all

areas was more positive than that of preceptors or residents, parent scores were lowest in areas that involved patient management, suggesting that residents were less effective at helping parents navigate the system of care than they were in their direct interactions with patients.

Chart-stimulated recall (see Chapter 8) is another method for assessing resident competence in SBP. Through detailed review of resident management as documented in the patient chart, along with queries to the resident about her reasons for those decisions, plans for additional follow-up, and ongoing illness prevention strategies, the faculty preceptor can gain insight into both resident knowledge and the application of that knowledge into the coordination of their patient's care.[7]

Self-reflection is a strategy through which a resident's self-assessment can enhance his appreciation for the barriers that often confront patients and families. Requesting residents to reflect on what they have learned from caring for a patient with a chronic condition who required multiple referrals and multiple sources of community support is another method for assessing both knowledge and resident understanding of how best to coordinate care for their patient. Self-reflection also provides an opportunity for residents to identify medical errors and systems problems that present barriers to effective patient management.

Finally, *peer assessment*, particularly of resident participation in group efforts to improve patient care systems, provides another perspective on a resident's engagement and effectiveness in identifying and correcting deficiencies in systems of care.

<u>We usually assess residents as individuals, but providing a medical home for patients requires a team approach to care. How can we understand the effectiveness with which they contribute to the efforts of a team to enhance patient safety and to improve care?</u>

Throughout residency training, in every setting, resident teamwork is an important part of effective patient care. Whether participating in resuscitation of a child in the pediatric emergency department or signing out patients to be cared for by colleagues overnight, residents depend on their colleagues to achieve the outcomes they want for their patients. In this era of limited work hours and reduced opportunities to participate in continuity clinic, patients are often seen in clinic by residents who are not their primary provider. In addition, the complexities of modern systems of care, whether in the outpatient or the inpatient arena, provide residents with multiple opportunities to work together on projects designed to enhance efficiency, communication, and cost-effectiveness. Faculty preceptors and resident colleagues are in the best position to observe and assess resident teamwork, and global assessment instruments with anchors describing a range of behaviors allow faculty and other residents to indicate the degree to which residents contribute to the functioning and leadership of teams.[8]

Continuity clinic-based projects initiated by residents and directed toward improving clinic functioning provide another opportunity for assessment. Delphin and Davidson[9] described a program for assessing team functioning in an anesthesia residency program. Self-selected teams of residents identified a project intended to enhance a system of patient care, and a faculty mentor, with the expectation that the project would be completed within a year and presented at a poster discussion session. The team was responsible for managing meetings and communication. Assessment of individual team member participation was requested annually from the other members of the team. Success of the projects was judged by the program director, department chair, and an outside expert in health care administration based on predetermined criteria, and degree of change in the organization as a result of the projects was used to evaluate the effectiveness of this program as an SBP educational effort. The authors describe successful implementation of projects that involved multiple disciplines and brought about change in their health care organization. They attribute the success and sustainability of their effort to the use of a resident team approach and

the expectation that faculty members assist and mentor resident teams. Similar resident-initiated projects could be stimulated by identification of errors or threats to patient safety.

How can advocacy for both patients and for improved systems be assessed?

Advocacy for improved care for patients and for improvements in the system in which their care is provided is an important part of every physician's responsibilities. Individual patient advocacy can be assessed by faculty as they precept residents and by nurses and other medical staff as they observe resident willingness to refer patients for services, reach out to schools and home care agencies on behalf of patients, and pursue follow-up of management plans. Advocacy for improved systems is demonstrated when residents follow through on their own suggestions for process improvement, report safety concerns, and engage in projects to enhance system functioning. Assessment of aims development, participation in measurement, and interpretation of outcomes of quality improvement projects provides a more formal means of determining the extent to which residents understand the elements of systems improvement. There are no validated standards for assessing these behaviors, but their success depends on knowledge of the functioning of the system as much as it does on enthusiasm and effort.

Lessons learned

- Assessment of resident competence in SBP depends on observation by faculty, peers, families, and medical staff who are aware of behavioral goals through faculty development efforts and assessment tools that describe a spectrum of engagement.
- Self-reflection and guided reflection through chart-stimulated recall can enhance the resident's sensitivity to the importance of the elements of this competency.
- Team-based projects allow residents not only to learn about system complexity but to disseminate system improvements.

References

1. Johnson JK, Miller SH, Horowitz SD. Systems-based practice: Improving the safety and quality of patient care by recognizing and improving the systems in which we work. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. *Advances in Patient Safety: New Directions and Alternative Approaches, Vol 2: Culture and Redesign*. AHRQ Publication No. 08-0034-2. Rockville, MD: Agency for Healthcare Research and Quality; August 2008. p. 321-330.
2. Accreditation Council for Graduate Medical Education. Outcome project—Practical implementation of the competencies. 2006. [Internet] http:// *www.acgme.org/outcome/*. Accessed May 18, 2011.
3. Medical Home Initiatives for Children With Special Needs Project Advisory Committee. The Medical Home. *Pediatrics* 2002; 110(1)184-186.
4. Kerfoot BP, Conlin PR, Travison T, McMahon GT. Web-based education in systems-based practice. *Archives of Internal Medicine* 2007; 167:361-366.
5. Brinkman WB, Geraghty SR, Lanphear BP, Khoury JC, Gonzales del Rey JA, DeWitt TG, Britto MT. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Archives of Pediatric & Adolescent Medicine* 2007; 161:44-49.
6. Zimmer KP, Solomon BS, Siberry GK, Serwint JR. Continuity-structured clinical observations: assessing the multiple-observer evaluation in a pediatric resident continuity clinic. *Pediatrics* 2008; 121:e1633-e1645.
7. Schipper S, Ross S. Structured teaching and assessment. *Canadian Family Physician* 2010;56;958-9.

8. Hammick M, Freeth D, Koppel I, Reeves S, Barr H. A best evidence systematic review of inter-professional education: BEME Guide No. 9. *Medical Teacher* 2007; 29:735-751.
9. Delphin E, Davidson M. Teaching and evaluating group competency in systems-based practice in anesthesiology. *Anesthesia and Analgesia* 2008; 106:1837-1843.

Annotated Bibliography

Ziegelstein RC, Fieback NH. "The mirror" and "the village": a new method for teaching practice-based learning and improvement and systems-based practice. *Academic Medicine* 2004; 79:83-88.

The authors describe implementation of daily multi-disciplinary inpatient rounds, monthly nursing evaluation of residents, and exercises to assess patient care quality as mechanisms to teach the concepts of systems-based practice. The metaphors of the mirror and the village were used to illustrate the concepts of practice-based learning and systems-based practice, respectively. This article describes introduction of these concepts and the success of the programs implemented for that purpose. It does not describe assessment of resident achievement of competence.

Dyne PL, Strauss RW, Rinnart S. Systems-based practice: the sixth core competency. *Academic Emergency Medicine* 2002; 9:1270-1277.

The authors describe the conclusions of a Consensus Group of the Council of Emergency Medicine Residency Directors (CORD) regarding appropriate assessment methodologies for the systems-based practice competency. They conclude that direct observation, global rating of observed performance by supervisors, 360-degree evaluation, portfolios describing projects and reflective exercises, standardized oral examinations, and written examinations to assess knowledge specific to SBP. They describe chart-stimulated recall, objective structured clinical examinations (OSCEs), and patient surveys as secondary assessment methodologies. The authors provide sample questions for assessment of SBP within the specialty of emergency medicine.

Hingle S, Rosher SB, Robinson S, McCann-Stone N, Todd C, Clark M. Development of an objective structured system-interaction examination. *Journal of Graduate Medical Education* 2009; 1:82-88.

The authors describe a variation on the objective structured clinical examination developed by faculty members at Southern Illinois University School of Medicine. They developed scenarios around 12 skills of systems-based practice and used standardized patients to illustrate the challenges of this competency. Both standardized patients and faculty members contributed to the feedback provided for residents.

Wittich, CM, Reed DA, McDonald FS, Varkey P, Beckman TJ. Perspective: transformative learning: a framework using critical learning to link the improvement competencies in graduate medical education. *Academic Medicine* 2010; 85:1790-1793.

The authors describe critical reflection on personal experiences with suboptimal patient care as a means of addressing both personal limitations and opportunities to improve the larger health care system.

Resources for Further Learning

The resources on these pages offer selected opportunities for learning more about assessment than can be covered in this primer.

Books

Brennan RL. *Educational Measurement*. Connecticut: Praeger Publishers; 2006.

Cizek GJ, Bunch MB. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage Publications; 2007.

Cooke M, Irby DM, O'Brien BC. *Educating Physicians: A Call for Reform of Medical School and Residency*. San Francisco, CA: Jossey-Bass; 2010.

Fitzpatrick JL, Sanders JR, Worthen BR. *Program Evaluation: Alternative Approaches and Practical Guidelines, 4th edition*. New Jersey: Pearson Education; 2011.

Joint Committee on Standards for Educational and Psychological Testing. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 1999.

Kern DE, Thomas PA, Hughes MT, editors. *Curriculum Development for Medical Education: A Six-Step Approach, 2nd edition*. Baltimore: The Johns Hopkins Press; 2009.

Yudkowsky R, Downing S, editors. *Assessment in Health Professions Education*. New York: Routledge; 2009.

Training Opportunities

The Accreditation Council for Graduate Medical Education (ACGME) Annual Educational Conference offers workshops and mini-courses focused on the needs of residency programs and program directors.

The Association of Pediatric Program Directors (APPD) Annual Meeting includes a variety of workshops on educational topics, including assessment.

The Academic Pediatric Association (APA) offers workshops at its conferences, including the Annual Pediatric Academic Societies meeting.

The Group on Educational Affairs (GEA) offers workshops at the Association of American Medical Colleges (AAMC) Annual Meeting as well as a series of medical education workshops at regional meetings through its Medical Education Research Certificate (MERC) program. Completion of six workshops entitles the learner to a MERC Certificate.

The Department of Medical Education at the University of Illinois at Chicago College of Medicine (UIC DME) offers a certificate program for program directors as well as a Master in Health Professions Education degree and a Ph.D. in curriculum studies with a concentration in the health professions. See *http://www.uic-dme.org* for more information.

The Division of Medical Education at the University of Southern California Keck School of Medicine offers fellowship programs in education leadership and teaching/learning, as well as a Master of Academ-

ic Medicine degree. See *http://keck.usc.edu/en/Education/Division_of_Medical_Education.aspx* for more information.

The Harvard Macy Institute offers a series of short courses in areas of medical education, including assessment. See *http://www.harvardmacy.org/* for more information.

Web sites

Western Michigan University hosts a web site devoted to best practices in the creation of evaluation checklists at *http://www.wmich.edu/evalctr/checklists/*.

MedEdPortal is a peer-reviewed resource for medical education materials hosted by the AAMC. Materials are available for free from *http://www.aamc.org/mededportal*.

APPD Share Warehouse, open to members of the APPD, hosts resources by and for pediatrics graduate medical educators at *http://www.appd.org/ed_res/share_warehouse.cfm*.

Glossary

360° assessment. See *multi-source assessment*.

Anchor. A numerical or descriptive label associated with a point on a rating scale.

Assessment instrument. The means through which an assessment method collects data to use as evidence of learning of knowledge, attitudes or skills.

Assessment method. A strategy, process or means by which one goes about gaining evidence about behavior(s) for the purpose of providing inference about a particular performance or achievement.

Assessment tool. See *assessment instrument*.

Authentic. Having *authenticity* (see below).

Authenticity. A property of an assessment method that refers to its use of natural settings and observation of naturalistic behaviors *in vivo*.

Blueprinting. Showing direct linkages between educational objectives and assessment contents. For example, on a multiple choice question formatted test, a test blueprint defines the proportion of test questions allocated to each topic area.

Calibration. Comparison of measurements by one method or individual to a standard, ostensibly "correct" measurement. In the medical education setting, calibration of performance ratings by individual observers involves achieving an understanding of expected goals for specific behaviors and assessing performance relative to those goals.

Case report poster. A poster presentation format that describes a clinical case intended to be informative to others about an interesting finding, presentation, diagnosis or management. Case reports are often used to prompt exploration of further understanding about a range of clinical topics.

Case specificity. A construct which indicates that physicians may not transfer knowledge, skills and attitudes learned in one clinical encounter or context to other cases/encounters or other contexts.

Central tendency. A consistent tendency of a rater to restrict their responses to the middle of the rating scale and to avoid ratings at one end or the other. Also known as "restriction of range" around the midpoint.

Chart-simulated recall. A standardized oral exam using examinees' patients' records. Allows the examiner to ask questions about clinical diagnosis and management based on actual patient records.

Checklist. Assessment items that are used to record whether intended behaviors were observed by an assessor. Typical checklist responses are dichotomous ("done" or "not done") but may use more categories (e.g. "done", "partially done", or "not done").

Compensatory scoring. An approach to portfolio assessment in which each component of the portfolio is scored separately and then averaged.

Competencies (ACGME). Six domains of learning that must be demonstrated by a trainee in order to graduate from a residency training program. The six domains are patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism and systems-based practice.

Competency-based medical education. An approach to training that focuses on whether competence is achieved in designated domains as the main outcome of learning.

Computer-based testing. An assessment method in which items are presented by computer, usually focused on measuring problem solving skills through patient management problems. Thought to offer greater authenticity than paper-based testing due to the ability to present multimedia items.

Conjunctive scoring. An approach to portfolio assessment in which each component of the portfolio must meet a minimum standard in order for the entire portfolio to be judged acceptable.

Construct underrepresentation. Inadequate or biased sampling of items in the area or domain of assessment. A threat to the validity of the assessment.

Constructed portfolio components. Components of an assessment portfolio that are chosen by the program director.

Construct-irrelevant variance. Systematic (non-random) error introduced into an assessment unrelated to the construct being measured that reduces the ability to meaningfully interpret scores or ratings. A threat to the validity of the assessment.

Context. The conditions, system, environment, and/or setting in which an activity or situation occurs.

Credibility. In qualitative research, an expression of the believability of the interpretation of the findings. Often assessed or enhanced by comparing multiple sources of evidence, expert review, and reflection by the subjects of the assessment.

Criterion-referenced. Test scores that are interpreted based on specific contents that residents or students actually know or can do. See also, for contrast, *norm-referenced*.

Critical incident. An incident of sufficient importance as to require reporting to a central authority or leader.

Critical incident technique. A method of research, evaluation, or quality improvement in which respondents generate and describe critical incidents in order to obtain insight on the process being studied.

Critically appraised topic. An evidence-based medicine activity in which learners a) identify an important clinical question; b) find and appraise the evidence available to inform the response to that question and related gaps in knowledge; and, c) synthesize and present their findings in written or oral format.

Cueing effect. The potential for a list of possible answers to influence an examinee's response to a multiple-choice question when compared with the answer that he/she would have given if the question had been asked in an open-ended format. Cueing can be positive (the learner does better because of the cue) or negative (the learner does worse because of the cue).

Cultural competence. The ability to interact effectively with people of different cultures.

Dependability. In qualitative research, an expression of the reproducibility of findings based on the use of consistent, documented research processes and external monitoring to ensure the processes are followed as documented.

Descriptive anchors. Scales that provide detailed descriptions of the performance expected at each score level as *anchors*. When descriptions are of behaviors, also called behavioral anchors.

Emotional intelligence. The ability to identify, understand, assess and manage the emotions of oneself and others.

Empathy. The capacity to recognize, understand, be sensitive to, and, to some extent, share feelings that are being experienced by another.

Entrustable professional activities (EPAs). A method of specifying and assessing clinical competence by means of defining units of everyday physician work (professional activities) and their observable outcomes, and determining the degree to which a learner can be entrusted to perform that work with the desired outcomes, at varying levels of supervision. Based on the work of Ten Cate and colleagues.

Faculty development. A process by which medical school faculty, including preceptors teaching in the clinical setting, participate in programs intended to improve their skills as educators, leaders, and scholars or investigators. Faculty development activities are successful when both the goals of the individual faculty member and the goals of the educational enterprise are being met.

Feasibility. The likelihood that an assessment is capable of being carried out and completed.

Feedback. Communication to an individual of their performance in relation to a standard of behavior or professional practice. Accurate and timely feedback has been shown to help learners improve their performance.

Formative assessment. An assessment that is used to inform the teacher and learner of what has been taught and learned, respectively, for the purpose of improving learning. Typically, the results of formative assessment are communicated through feedback to the learner. Formative assessments are not used to make judgments or decisions. See, in contrast, *summative assessment*.

Global assessment/rating. Rating scales that rate performance as an integrated whole. For example, "Overall this performance was: excellent, very good, good, marginal, unsatisfactory".

Halo effect. A tendency of raters to allow their perception of an individual's performance in one trait to influence their rating of performance in other traits.

Hawthorne effect. The tendency of the act of observation to change the behaviors of those observed.

High stakes assessment. A measurement of learning that is used for making a decision of significant consequence. The American Board of Pediatrics Certifying examination is an example of a high stakes assessment.

Humanism. A focus on human values and concerns and the duty to promote human welfare.

Identity development. A process by which a learner matures in thought and emotion, moving from an egocentric focus, to one that appreciates, reconciles, and values the perspective of others.

In-training examination. A single-best-answer multiple choice format examination administered yearly by the American Board of Pediatrics to all pediatrics residents in the United States to assess medical knowledge.

Key feature test. A testing format that focuses on critical decision making in clinical settings. It is based on the concept that in any clinical encounter there are a small number of essential decisions that form the key steps or key features in the successful resolution of the problem.

Leniency. A tendency of a rater to give learners higher ratings than they should receive based on their actual performance.

Likert scale. A scale format commonly used in questionnaires in which the respondent expresses or rates their level of agreement with statements. Traditionally, there are five different levels of agreement. A variety of similar scales with different numbers of levels or anchors are sometimes referred to as "Likert-type scales". Although "Likert" is pronounced variously, Rensis Likert, the originator of the format, pronounced the first syllable of his last name with a short "i" vowel sound (like "lick", not "like").

Metacognition. Knowledge, awareness and understanding about one's own thinking.

Milestones. Points along a developmental continuum that mark the accomplishment and integration of specific knowledge, skills and attitudes that allow one to perform at particular levels. May also refer to a joint initiative of the ACGME and ABP to define and refine the ACGME competencies in the contest of pediatrics.

Miscalibration. A failure of one measurement to predict another measurement. For example, low accuracy in self-assessment (underestimation or overestimation of one's abilities) suggests that self-assessments will be miscalibrated with respect to assessments by observers.

Multiple-choice question. A test question in which the examinee is given a question followed by a list of several possible answers to choose between. Depending on the format, one or more than one answer may be correct.

Multi-source assessment. The use of assessments made by multiple people who work with a learner (often in different capacities) in order to form a more global assessment of the learner. Also known as 360° assessment (especially when individuals are evaluated by their superiors, subordinates, peers, and self) or multi-rater feedback.

Needs assessment. A process for determining needs, or "gaps" between current conditions and desired conditions. In the context of medical education, a needs assessment might investigate the gap between the curricular and assessment elements of the current program compared to the educational goals for that program, or the gap between what learners currently know and what they should know in order to practice.

Norm-referenced. Test scores that are interpreted relative to some well-defined normative group. A normative group could be the group of residents who took the test or assessment. See also, for contrast, *criterion-referenced*.

Objective structured clinical exam (OSCE). An assessment format that consists of a series of performance tests. Each test within an OSCE is called a "station".

Outcome Evaluation. Evaluation that measures achievement of intended outcomes by a program.

Patient management problem. An assessment item format that begin with a patient's presenting complaint. The examinee is asked to select appropriate items of history, examination, and investigation before making a diagnosis and outlining a management plan.

Peer Assessment. Assessments of a learner performed by their peers (as distinct from their teachers, themselves, etc.)

Portfolio. In assessment, a collection of evidence of progression towards proficiency (e.g. in the ACGME competencies). Portfolios typically include both constructed components (selected by the program or faculty) and unconstructed components (selected by the learner).

Process Evaluation. An evaluation method that focuses on how a program is implemented or operates. It identifies procedures and processes such as decision making.

Professional conduct. Behavior that is recognized as acceptable according to standards that are set by professionals within a group.

Professional formation. An experiential process by which one learns to take on the identity and values of a physician, typically through observation and modeling of other physicians.

Professionalization. A social process whereby a person takes on the characteristics and culture associated with a professional role.

Program Evaluation. A systematic method of collecting, analyzing and using information to answer basic questions about the activities and outcomes of a program or project.

Quality improvement project. A learning activity that aims to improve healthcare quality, often using a sequence of constructed steps designed to instruct and assess knowledge, critical thinking, and understanding of healthcare systems.

Reflection-in-action**.** A task-bound reflective process in which one continues to act but reshapes one's actions in real time through explicit cognition. It is a dynamic and ongoing monitoring process. Associated with the work of Donald Schön. See, in contrast, *reflection-on-action*.

Reflection-on-action. The reflection that occurs following an event or experience and incorporates one's current knowledge of a situation or problem and addresses how the situation could have been handled differently. Associated with the work of Donald Schön. See, in contrast, *reflection-in-action.*

Reflective practice**.** The ability to critically think about and examine one's own reasoning and decisions in order to improve judgment and develop expertise. Associated with the work of Donald Schön.

Reliability. The degree to which an assessment can be replicated or reproduced. There are multiple approaches to the measurement of reliability with different associated statistical methods. Reliability may consider, for example, whether the same assessment would yield similar results if conducted by multiple raters or over two points in time.

Script concordance test. An assessment format that attempts to measure the organization of clinical knowledge in the mind of the examinee. Script concordance questions present an examinee with a clinical scenario and provide new elements of information in a stepwise fashion. Grading of the question is ac-

complished by comparing the similarity of the responses of the examinee with those of a panel of experts presented with the identical scenario.

Selected-response format. A test question in which the examinee is asked to select a response from a list of answers provided as opposed to an open-ended (or constructed-response) question such as an essay in which the examinee is asked to generate a response on his/her own.

Self-assessment. An individual's evaluation of their own performance, skills, attributes, or abilities.

Self-concept. A broad cognitive appraisal of oneself formed by external feedback and introspection. Self-concept influences self-assessment.

Self-efficacy. A context-specific assessment of competence to perform a specific task. It is an individual's judgment of his/her capabilities to achieve a given goal.

Severity. A tendency of a rater to give learners lower ratings than are warranted based on their actual performance.

Stakeholder. A person or group that has an investment, share or interest (direct or indirect) in an educational program.

Standardized or simulated patients. Actors trained to play the roles of patients, portray specific cases, and rate performance of the learner. Often used in OSCE stations.

Summative assessment. Evidence of learning that describes the composite performance of the development of a learner at a particular point in time. Summative assessments may be used to make decisions or judgments about the level of learning required for a particular score, grade or other criteria. See, in contrast, *formative assessment*.

Triangulation. In qualitative research, the use of evidence from multiple sources or perspectives to lend credibility to a conclusion. Assessing a resident from multiple perspectives is an example of triangulation.

Unconstructed portfolio components. Learner-selected items that are included in a portfolio. For example, unconstructed components may include best work products (e.g., presentations, critically appraised topics, etc.) and reflections by the learner.

Utility. In the context of assessment, the concept (associated with van der Vleuten) that the usefulness of an assessment method is based on a combination of reliability, validity, educational impact, acceptability, and cost.

Validity. The evidence presented to support the meaning assigned to assessment data. Validity is the degree to which an assessment is "measuring" what it is supposed to measure.

About the Contributors

Ann E. Burke, MD is Associate Professor at Wright State University Boonshoft School of Medicine in Dayton, Ohio. After graduation from University of Virginia School of Medicine in Charlottesville, VA, Dr. Burke began residency on active duty with the United State Air Force at the Wright State University Integrated Program at Dayton Children's Medical Center and Wright Patterson Air Force Base (WPAFB). After finishing her commitment to the Air Force, she became the Pediatric Program Director at Wright State and has been there for the past 11 years. She currently serves as President of the Association of Pediatric Program Directors. Dr. Burke is a member of the Milestones Working group, the American Board of Pediatrics Program Director's Committee and FOPO Task Force on Women in Medicine. She is married to an internist, Brian Burke and has three wonderful children.

Carol L. Carraccio, MD, MA recently assumed a position at the American Board of Pediatrics as Director of Competency-based Assessment Programs. Prior to this she served as the Program Director/Associate Program Director at the University of Maryland for twenty six years. She has served in a number of national leadership roles over the course of her career including President of the Association of Pediatric Program Directors, Chair of the Pediatrics Residency Review Committee, Chair of the Program Director's Committee of the American Board of Pediatrics, Director of the Initiative for Innovation in Pediatric Education and Chair of the Pediatric Milestones Working Group. Her research and scholarly work has focused on competency-based medical education and assessment, with a particular interest in portfolio assessment.

Patricia J. Hicks, MD, is Professor of Clinical Pediatrics in the Department of Pediatrics in the School of Medicine at the University of Pennsylvania School of Medicine and the Director of the Pediatric Residency Program at The Children's Hospital of Philadelphia. She is the president-elect of the Association of Pediatric Program Directors (APPD) and Chair of the APPD Longitudinal Educational Assessment Research Network (LEARN) Advisory Board. She is a member of the Initiative for Innovation in Pediatric Education (IIPE) Review Committee, a member of the Pediatrics Milestones Working Group and a member of the Program Directors Committee of the American Board of Pediatrics.

M. Douglas Jones, Jr., MD is Professor of Pediatrics and Senior Associate Dean for Clinical Affairs at the University of Colorado School of Medicine. He was Chair of the Department of Pediatrics at the University of Colorado School of Medicine and L. Joseph Butterfield Chair in Pediatrics and Pediatrician-in-Chief at The Children's Hospital from 1990-2005. He is Immediate Past-Chair of the Board of Directors of the American Board of Pediatrics. He chaired the Residency Review and Redesign in Pediatrics (R3P) project along with Dr. Gail McGuinness. He served on the Review Committee for Pediatrics of the Accreditation Council for Graduate Medical Education from 1999-2005 and as Chair in 2004-2005. He has received the Kempe Professional Service Award, the High Hopes Award from the Children's Diabetes Foundation and the Joseph W. St. Geme, Jr. Leadership Award from the Federation of Pediatric Organizations.

Stephen Ludwig, MD is Professor of Pediatrics and Emergency Medicine at The University of Pennsylvania School of Medicine. He serves as the Designated Institutional Official for The Children's Hospital of Philadelphia where he has worked for the last thirty-seven years. For fifteen years he was the Program Director of the Pediatric Residency. Dr. Ludwig is on the Board of Directors of the American Board of Pediatrics. He is the Chair of the Pediatric Review Committee of the Accreditation Council for Graduate Medical Education. He was past-president of the Academic Pediatric Association and was elected to the Institute of Medicine.

Gail A. McGuinness, MD is Executive Vice President of the American Board of Pediatrics (ABP) as well as a Director of the ABP and the ABP Foundation. She serves as the ABP's liaison to the Review Committee (RC) for Pediatrics of the Accreditation Council for Graduate Medical Education (ACGME), as well as a liaison to the Committee on Pediatric Education and the Committee on Pediatric Workforce of the American Academy of Pediatrics. She has been elected to serve on the Board of Directors of the American Board of Medical Specialties (ABMS). She holds appointments as a Clinical Professor of Pediatrics at Duke University Medical Center and at the University of North Carolina-Chapel Hill. Dr. McGuinness previously held a position as Professor of Pediatrics and Residency Program Director at the University of Iowa, where she also served as Associate Chair for Education in the Department of Pediatrics.

Julia A. McMillan, MD, is Vice Chair for Education and Residency Program Director in the Department of Pediatrics at John Hopkins School of Medicine. She also serves as Associate Dean for Graduate Medical Education for the School of Medicine. She is a member of the Residency Review Committee for Pediatrics of the Accreditation Committee for Graduate Medical Education and a member of the Program Directors Committee of the American Board of Pediatrics.

Alan Schwartz, PhD, is Associate Professor and Director of Research in the Department of Medical Education and Associate Professor in the Department of Pediatrics at the University of Illinois at Chicago. He serves as a member of the project support team for the Initiative for Innovation in Pediatric Education (IIPE) and senior consultant to APPD's Longitudinal Educational Assessment Research Network (LEARN). His primary area of research is medical decision making. He received the Ray E. Helfer Award for Innovation in Pediatric Education from the Ambulatory Pediatrics Association in 2001 for research on assessment of evidence-based medicine skills.

Richard P. Shugerman, MD is Professor of Pediatrics and Director of the Pediatric Residency at the University of Washington School of Medicine/Seattle Children's Hospital. He works clinically in the emergency department at Seattle Children's Hospital. His primary areas of research are medical education and physician career satisfaction. He received the Parker J. Palmer "Courage to Teach Award" from the Accreditation Council for Graduate Medical Education in 2007 and the Robert Holm Award for leadership in residency education from the Association of Pediatric Program Directors in 2009.

Suzanne K. Woods, MD is Associate Professor of Pediatrics and Internal Medicine at Duke University Medical Center. She is the combined Medicine-Pediatrics program director, chief of the Medicine-Pediatrics faculty section and Med-Peds clinic director. She has served on the Medicine Pediatrics Program Directors Association (MPPDA) executive committee and is the past president and leader of the Accreditation committee. Her clinical and educational interests include vaccines, transitional care, and development of a resident advising program and individualized learning plans.