AAIM
Alliance for Academic Internal Medicine
Association of Professors of Medicine
Association of Program Directors in Internal Medicine
Association of Specialty Professors
Clerkship Directors in Internal Medicine
Administrators of Internal Medicine

AAIM Perspectives

AAIM is the largest academically focused specialty organization representing departments of internal medicine at medical schools and teaching hospitals in the United States and Canada. As a consortium of five organizations, AAIM represents department chairs and chiefs; clerkship, residency, and fellowship program directors; division chiefs; and academic and business administrators as well as other faculty and staff in departments of internal medicine and their divisions.

Direct Observation of Residents: A Model for an Assessment System

Anthony A. Donato, MD
Internal Medicine, Reading Health System, West Reading, Pa.

Consider the following vignette:

You observe your second-year resident physician admit a late-stage dementia patient suffering from his fifth pneumonia in the past 12 months. During the interview, the patient's wife says to your resident, "He has such a poor quality of life. I don't think he'd want to go on like this." The resident deflects the comment, telling the wife that he is confident that the team will be able to make the patient's pneumonia better. When you ask him afterward about that comment, he tells you that he is not comfortable with end-of-life issues in patients but says he was "going to consult the palliative team anyway."

How can this observation be best used to maximize possibly competing purposes: the formative effect on the resident's learning, the summative decisions made about the resident's competence, and the effectiveness of the educational program and assessment system overall?

Graduate medical education in North America and Europe is in the midst of a rapid and significant transformation. Time-based curricular designs, in which demonstration of medical knowledge was paramount, are now being replaced by outcomes-based designs, whereby resident performance in actual practice is the primary measured outcome.[1,2] Capturing the complex behaviors defined in the competencies and milestones as defined by national organizations during direct observation of residents will require assessment tools that can record rich, meaningful narratives of performance of directly observed resident behaviors.[3] Deliberately constructed systems of assessment will be critical to programs that are now faced with managing large volumes of narrative accounts of resident performances, in order to efficiently collect, aggregate, value, and use this information for optimal formative and summative assessment decisions, as well as for feedback to the assessment system itself.[4]

The proposed model is an attempt to integrate the performance data obtained from direct observations of residents into a larger assessment system. We will begin by discussing theoretical issues related to optimal use of direct observation with regard to formative and summative purposes of the residents, as well as system feedback to the assessment system. We then will discuss the interrelations of resident, mentor, summative evaluator, curriculum, and observer as they pertain to the various sources and types of information from a direct observation system, and will apply the model to the introductory vignette presented above.

ISSUES TO CONSIDER REGARDING FORMATIVE FEEDBACK TO INDIVIDUAL RESIDENTS

Any assessment action will result in an educational reaction from learners.[5] Given that the ultimate goal of training programs is to promote effective learning,

Funding: None.
Conflict of Interest: None.
Authorship: The author attests that he had access to all of the data and was involved in the planning and writing of the manuscript.

Requests for reprints should be addressed to Anthony A. Donato, MD, Department of Medicine, B-2, Jefferson Medical College, 6th and Spruce Streets, West Reading, PA 19612.
E-mail address: donatoa@readinghospital.org

http://dx.doi.org/10.1016/j.amjmed.2014.01.016

a program's assessment systems must be designed with the intent of maximizing this educational impact. Assessment instruments should foster meaningful, timely, specific, task-focused, and goal-oriented feedback.[6] Offering rating scores along with narrative comments blunts the educational effects of those comments.[6] De-prioritizing scoring aspects of direct observation may prevent scores from inhibiting learning.[6] Any single performance observation is context-specific,[7] and therefore, too limited for any meaningful interpretation of overall competence. However, aggregate narratives of performance from multiple sources will help the learner to calibrate their overall progress toward learning goals. Assessment instruments should therefore facilitate the recording of rich narratives of direct observations as well as written action plans for meaningful use by the resident (reflection-*on*-action).[8] Because self-assessments often poorly correlate with external measures of performance ($r = 0.3$), especially for those at the extremes of performance,[9] residents will benefit from comparing their self-assessments with those from multiple independent observers performing direct observations. Structured questions in a personal development portfolio may be a useful adjunct to support self-reflection.[10] Added guidance by a dedicated physician mentor can further enhance a resident's development of personal improvement by sharing and discussing portfolio reflections.[11,12] To avoid conflicting purposes and compromising the authenticity of their learning facilitator role,[13] the mentoring physician should limit his/her role to a resident advocate only, and should not be part of the summative decision-making process.

PERSPECTIVES VIEWPOINTS

- Direct observation will play a critical role in competency-based curricular designs.
- Development and implementation of systems of assessment will be critical to programs in order to efficiently collect, aggregate, value, and use this information for optimal formative and summative assessment decisions, as well as for feedback to the assessment system itself.

ISSUES TO CONSIDER REGARDING SUMMATIVE ASSESSMENT

Any single assessment point, from a single question on a medical knowledge test, a single station on an objective structured clinical examination, or a single direct observation, is highly context-dependent and therefore, too limited as the sole basis for any high-stakes assessment.[3] However, a robust summative assessment of competence can be reached by a program director and competence committee by collating expert judgments from multiple sources, and by applying rigorous qualitative methods (triangulation, prolonged engagement, and member checking) to create a case for competency.[3] To overcome the limitations of case-specificity and unwanted observer rating variance, broad sampling, using multiple independent observers and cases, is necessary.[7] Therefore, assessment tools should be versatile enough to sample from all venues in which residents practice. Also, because they represent important sources of curricular evaluation, the information gathered from direct observations can provide a measure of program effectiveness, especially if the information is electronically recorded and collated by topic area (eg, "geriatrics") and skill type (eg, "oral presentation").

ISSUES REGARDING FEEDBACK TO THE ASSESSMENT SYSTEM

Faculty observers are "active reasoners" in the process of making observations, using the same problem-solving skills and scripts that they use in patient care.[14] Observers also make spontaneous, idiosyncratic "social judgments," using 2 orthogonal dimensions that are often categorized as "socially desirable traits" (eg, "communication" or "warmth") and traits that affect an individual's success (eg, "knowledge" or "intellect").[15] Observers have difficulty translating these holistic impressions into multidimensional numerical scales.[15,16] Efforts to add structural changes to rating forms (eg, anchors, scale lengths) to decrease observer variance have had little impact.[17] This has led some to state that "validity resides in the users, not the instruments."[3] Employing faculty observers to choose elements for assessment tools that match their goals and priorities may be an important strategy to gain meaningful direct observation data.[18]

Observers also are learners themselves.[3] They must learn to simultaneously diagnose patients and residents, which may overwhelm their observational capacity. More senior observers develop heuristics ("performance scripts") to process this information and increase observational capacity.[19] Aligning assessment instruments with observer tasks may reduce extraneous cognitive load associated with use of instrument itself. Feedback to observers from the assessment system with regards to their performance as observers may help them develop performance scripts and become more effective observers.

In accordance with deliberate practice theory, all learners should be placed in continually more challenging environments as their skills accrue.[20] An effective assessment system should facilitate this process, identifying skills a resident has mastered as well as areas of weakness.[3] This can allow the program to re-direct future observations away from a venue they may have mastered (ie, "handoffs") and toward an area of need (eg, "communication skills in difficult family meetings").