Instructors guide

Development and Implementation of Multi-source Assessment Tools for ACGME Residents and Fellows

Key Words- assessment, evaluation, multi-source, 360 degree, GME, resident, fellow, core competencies, feedback, milestones, Next Accreditation System (NAS)

Educational Objectives

1. Utilize formative competency-based, multi-source evaluation of residents and fellows in ACGME- accredited Graduate Medical Education (GME) programs to inform summative milestone assessment
2. Provide formative multi-source assessment tools to ACGME programs not yet reporting milestones
3. Provide multi-source assessment tools to non-ACGME programs in an effort to standardize an approach to multi-source evaluation
4. Promote the use of multi-source evaluation tools as the cornerstone for end-of-rotation written evaluation, and as a vehicle for facilitating verbal feedback
5. Provide standard templates for multi-source trainee evaluations that can be used across specialties and customized, as needed
6. Utilize an institutionally-approved assessment tool as a key element in faculty development focused on assessment
7. Create an opportunity to benchmark performance and facilitate comparisons across programs and specialties

Resource Description

Background

Effective evaluation is a cornerstone of medical education. National data from the ACGME's 2013 resident survey indicates dissatisfaction with faculty feedback among 27% of residents; this finding has been consistent over several years (1). ACGME migration to milestones as part of the Next Accreditation System (NAS) has helped to focus increased GME Program Director and faculty attention on assessment and underscores the potential usefulness of thoughtfully developed standardized evaluations. Semi-annual reporting of individual achievement of

specified milestones must rely upon ongoing formative (e.g. end of rotations) assessment utilizing a variety of tools. Indeed the ACGME continues to require formative, competency-based, multi-source assessment (section V.A.2 of the Common Program Requirements 2013) at the conclusion of each rotation or educational assignment

In 2011, the Partners HealthCare System (including Massachusetts General and Brigham and Women's hospitals) appointed a 16-member education subcommittee to enhance evaluation and feedback in the system's 100+ accredited GME programs. The *Evaluation and Feedback Subcommittee* (EFS) included program directors and other educators from multiple specialties, institutional GME leadership, and trainees.

Ratings of evaluation and feedback on trainee surveys were suboptimal, and a high number of evaluation-related citations had been received as part of accreditation letters. We noted that the content and quality of assessment tools in our institutions were highly variable: evaluation forms used a variety of rating scales, often without anchors and lacking clear definitions and instructions; questions were frequently double- or triple-barreled, leading to confusion for both evaluators and trainees; prompts or requirements for text comments were often lacking. In addition, supervising faculty often failed to complete written evaluations as requested. Moreover, only a small minority of our accredited programs were utilizing multi-source evaluation as was newly required by ACGME.

An Education Specialist (physician with formal training in medical education) was hired to help lead and implement this effort. The EFS analyzed the barriers to achieving consistent, effective and timely evaluation and feedback. The group identified *development of template multi-source evaluation tools as a key resource need. Faculty development* and an *electronic platform for managing distribution and collection of evaluations* were also highlighted as important needs. In an effort to elevate the quality, consistency, and effectiveness of this evaluation and feedback initiative, the Education Specialist attended a one week, American Board of Internal Medicine (ABIM) Faculty Development Course, "Evaluation of Clinical Competence." In May 2011, the Partners Office of Graduate Medical Education invited ABIM leaders, Drs. Eric Holmboe and Bill Iobst, to speak at the "kick-off" faculty development event which featured a two day retreat for Program Directors at Partners Healthcare, *"Evaluation of Clinical Competence: Assessment and Evaluation Skills for Faculty in a New Era."* Program Directors from surrounding academic medical centers in the greater Boston area were also invited to attend.

The principle goal of developing common template evaluation forms was to elevate the quality of evaluation forms in use across our programs in order to enhance the quality of evaluation itself. Additional anticipated advantages were a) ease in aggregating and interpreting data for trainees rotating on specialty services outside of their primary program; and b) system-wide trend analysis relating to individual core competencies (such as systems-based practice). In light of a subsequent phased implementation of milestones for individual specialties, we expected that use of these forms would immediately elevate the quality of evaluation (especially among programs currently using particularly poor forms), and then over time items from the evaluation forms could be mapped to specialty-specific milestones to inform Clinical Competency Committees in assessing achievement of milestones.

Method for Developing the Tools

The Education Specialist presented results of a literature review and of interviews with educational leaders from four major teaching institutions across the country regarding their approach to multi-source evaluation of GME trainees. In addition, a number of assessment tools from various Partners training programs and other teaching institutions were examined. The EFS concluded that ideal, competency-based 360-degree assessment tools were not generally available. Thus, the group embarked on a consensus-based process to develop these tools, informed by the literature and by identified "best practices".

The following considerations were the focus of thoughtful discussion and debate:

- *Content* – which specific components of the core competencies and which other elements of resident/fellow performance are relevant to assess and which could be effectively evaluated by each type of evaluator.
- *Evaluators* – which role groups would have sufficient contact with residents/fellows and sufficient context to effectively evaluate performance.
- *Number of items* - attempting to cover a sufficiently broad range of competencies without being unwieldy.
- *Type of scale* – i.e. relating to "expectations" for level of training, achievement (level of competency, such as according to Dreyfus model), or based on imputed comparison to others in the program or nationwide ("Superior", "Above Average", "Average", etc.).
- *Granularity of ratings* – i.e. the number of gradations (3-point scale, 9-point scale, etc).
- *Use of anchors* – descriptive text to illustrate of what is meant by various ratings.
- *Requirement for text comments*

We endeavored to develop evaluations that met the following criteria:

- Sufficient length and detail to capture essential information, without being so long or detailed so as to deter evaluators from completing them;
- Clear and easily understandable questions or prompts that would facilitate a common understand between evaluator and trainee
- A combination of ratings and text comments, based on the belief that each element serves a useful purpose.
- Incorporation of screening questions that would alert program directors to potential concerns.

Key decisions were made after extensive debate and consultation with stakeholder groups outside of the committee. Committee members were asked to review the evaluations and provide comments as to readability, length, and content. The system-side education committee then endorsed the resulting assessment tools. Our programs were presented with the opportunity to review and adopt or adapt these evaluations but could also decline to use them.

Results

1) Assessment tools.

The assessment tools reflect the following key decisions of the EFS:

- Five evaluator roles were identified, and a template evaluation form was developed for each: Self (resident or fellow); Peer; Faculty; Other health professional (e.g. nurse); Patient. [Note: Regulatory barriers to patient surveys regarding resident performance were encountered; these will be described and the tool will be published separately.]

- Content included relates to knowledge, attitudes and behavior relevant to all specialties. Of note, five questions that indicate concern in the areas of patient safety, ethics, integrity, and professionalism (4) were included to facilitate early identification of "at-risk" trainees.

- "Short" and "standard" versions of the faculty assessment tool were developed so that program directors could individually select the level of granularity to pursue. Additional customization is facilitated by provision of an appendix with 76 additional competency-based questions that can be easily incorporated. The intent of the appendix is to offer a series of questions that could provide a more granular assessment of the six competencies within the realm of specialty and sub-specialty medicine while retaining all or most of the common elements including in the template. (So far, programs have not chosen to customize the forms.) Programs are advised that customization of evaluations may impact the validity of the assessment tool.

- No items were designated as mandatory for the initial implementation; however, to the best of our knowledge, programs are using the assessment tools "as is" with all items included for trainee assessment. The group settled on a 4 point rating scale utilizing an "expectation" framework: Unsatisfactory, Below Expectations, Meets Expectations, Exceeds Expectations

- Text comments were considered essential, and some had argued to include text comments without quantitative ratings. It was decided that text comments would be required for all questions in which the trainee received an unsatisfactory rating

- The resulting assessment tools provide a means to gather formative multi-source, competency-based evaluation for assessment of ACGME and non-ACGME trainees. They have been provided to the approximately 100 ACGME-accredited GME programs at Massachusetts General and Brigham and Women's hospitals and are in use by a growing number of them—currently 36 ACGME programs utilize these evaluation forms including 33 specialties and sub-specialties. A substantial number of ACGME trainees are being assessed by these evaluations; currently, Faculty Evaluation of Trainee- short or standard (489 trainees), Evaluation of Trainee by Other Healthcare Professional (302 trainees), Peer evaluation of trainee (92 trainees), and Trainee self-assessment (196 trainees). In addition, 20 non-ACGME programs at Massachusetts General and Brigham and Women's hospitals are using the evaluations. The Partners Office of Graduate Medical Education works closely with GME programs and are not aware of any difficulties implementing the evaluation forms.

Validity:

Validation of the multi-source assessment tools was achieved using both a traditional psychometric and a contemporary validity framework, similar to the validation of the milestones (3, 10). Literature on complex performance assessments, (e.g. multi-source evaluations and the milestones), supports the use of contemporary validity approaches as measures of reliability or validity, which "should be combined with other quality criteria that are especially important for competency assessment" (3).

The traditional components of validity for these assessment tools include: Construct validity which incorporates the following features: Content validity, response process, consequence and content quality and coverage (5, 10). The contemporary components of validity for these assessment tools include: Meaningfulness, authenticity, fairness, comparability, acceptability, and transparency (3, 10).

<u>Traditional Components of Validity:</u>

Content validity: The evaluation tools were designed to assess trainee performance in the ACGME six core competencies as described in the Common Program Requirements (CPR) 2013. These core competencies are well known to trainees and faculty as they form the foundation of both the educational curriculum and evaluation process in accredited training programs. In the ACGME 2013 Common Program Requirements (2), program faculty (II.B. 1.b.) are to "administer and maintain an educational environment conducive to educating residents in each of the ACGME competency areas."

The 16 member multi-specialty EFS working group (composed of Program Directors, GME educators, GME leadership, and trainee representatives) developed a consensus on how the competencies could be evaluated across multi-specialty training programs. This process was

informed by knowledge of specialty curricula and their relationship to the ACGME core competencies, existing ACGME standards, and a review of the literature. We also examined available evaluations from other institutions as well as those within our own system.

Content quality and coverage: The content quality and coverage of the assessment tools are informed by relevant literature and review of other specialty and generalist multi-source evaluations from both within and outside the Partners Healthcare System.

Response process is aided by using the existing ACGME framework of the six core competencies, which form the cornerstone of accredited training program curricula and serve to anchor the multi-source evaluation tools. In our institution, the roll-out of new template evaluation forms created an opportunity to re-focus faculty (and other evaluators) on the importance of evaluation and feedback, and to provide training aimed at enhancing this process. A series of faculty and trainee development sessions have been ongoing since 2011. To date, over 800 faculty and trainees from the Partners system have attended an interactive conference or Grand Rounds utilizing video to "practice" using the ACGME six core competencies (and Milestones, if available) as a framework for trainee feedback and evaluation. These sessions have contributed to educating faculty and trainees on the assessment tools and process. Participants are encouraged to ask questions or request clarification.

These sessions are provided in a variety of forums including: Specialty-specific conferences, retreats, and grand rounds; Partners-wide GME programs and workshops and Partners Healthcare GME- sponsored clinical intern and fellow retreats. The specific items of the "Faculty Evaluation of Trainee" template are reviewed in each interactive conference. At the conclusion of the program, all attendees are given the opportunity to provide feedback and to rate how well the program met the following objectives:

1) Understand how their specialty training goals and objectives and the ACGME core competencies (and Milestones, if available) provide the cornerstone for evaluation and feedback;

2) Provide strategies for structured observation in clinical settings;

3) Understand the difference between feedback and evaluation;

4) Understand the components of meaningful feedback.

The rating scale for the workshop evaluation is "Strongly agree", "Agree", "Disagree". Participants are also encouraged to provide written comments on program strengths and suggested areas for improvement. Overall, the conference feedback has been very positive and indicates that the majority of attendees "Strongly agree" or "Agree" that the objectives have been met.

Faculty who attend these workshops are usually eligible to receive Continuing Medical Education credit through their departments, issued either by Harvard Medical School or the Partners Office of Continuing Medical Education.

The resulting assessment tools provide formative multi-source, competency-based evaluation for assessment of ACGME trainees. They have been provided to the approximately 100 ACGME-accredited GME programs at Massachusetts General and Brigham and Women's hospitals and are in use by a growing number of them—currently 36 ACGME programs utilize these

evaluation forms including 33 specialties and sub-specialties. A substantial number of trainees are being assessed by these evaluations; currently, Faculty Evaluation of Trainee- short or standard (489 trainees), Evaluation of Trainee by Other Healthcare Professional (302 trainees), Peer evaluation of trainee (92 trainees), and Trainee self- assessment (196 trainees).

Consequence:
These assessment tools have a role in both summative and formation evaluation. In terms of the summative component, all the assessment tools, with the exception of 'Trainee Self-Assessment,' are intended for end of rotation written evaluation as part of the competency-based multi-source formative evaluation process required by the ACGME. In our programs, and as far as we know regarding other programs, rotation evaluations (particularly in the aggregate) are an important component—along with other sources of information—that informs Program Director decisions regarding resident advancement. We believe that the process for developing these evaluation forms was considerably more deliberative than typically undertaken by Program Directors. The evaluation tools contribute to formative evaluation by providing framework for verbal feedback during the rotation.

Contemporary components of validity:
The contemporary components of validity for these assessment tools include: Fitness of purpose, meaningfulness, authenticity, fairness, comparability, acceptability, and transparency.

Fitness of Purpose (3) is considered the foundation for the development of all competency assessment programs (CAP) which prescribes that "all CAPs must be aligned with the goal of the learning process (i.e. the acquisition of competencies) and with the instruction given." The multi-source assessment tools align with the ACGME core competencies, inform summative milestone assessment, and are easily understood and implemented.

Meaningfulness:
The items in the multi-source assessments are based on the six ACGME core competencies which in turn, inform the competency-based milestones. Trainees aspire to achieve milestone progression and thus a formative competency-based multi-source assessment should increase meaningfulness for both the trainee and faculty (3 and 10). Meaningfulness could be studied by surveying GME training programs (e.g. trainees and faculty) that use these competency based formative multi-source evaluations and those who do not to determine trainee/faculty understanding of the six ACGME core competencies, their relationship to specialty-specific milestones, and their respective importance in informing formative vs. summative feedback.

Authenticity: Authenticity (3) "relates to the degree that a complex performance assessment resembles future professional life and assesses the competencies needed in the future workplace". The multi-source formative assessments are directly related to the ACGME core competencies which provide the framework for summative milestone assessments and the

roadmap to competency acquisition in the trainees' chosen specialty. As more specialties incorporate semiannual summative milestone assessments per ACGME requirements, reassessment of newly validated milestones and their relationship to ACGME core competencies will be important.

Fairness*:* The multi-source evaluation tools do not appear to show bias to particular group(s) of learners, rather, they reflect the knowledge, skills, and attitudes of the ACGME competencies they were designed to assess. This could be tested by a study that evaluates assessment data on a cohort of United States trainees who graduated from U.S. medical schools compared to assessment data from a cohort of residents/fellows training in the U.S. after graduating from international medical schools."

Comparability: Through faculty development, we have sought to teach program directors, faculty, and trainees about the assessment tools in an effort to have them utilized in a consistent manner.

Acceptability: The content of the multi-source assessment tools and the foundation upon which they were developed are accepted by a large number of multi-specialty program directors and educators who have implemented them.

Transparency: Through extensive ongoing faculty development in our large academic healthcare system and trainee sessions on feedback and evaluation, we have made an effort to ensure the multi-source assessment tools are clear and understandable to both evaluators and to those being evaluated. Faculty and trainees are educated on rating scales, the different evaluator groups completing the assessment (per ACGME multi-source evaluation requirements), and the purpose of the evaluation(s). Programs that use these multi-source evaluations are encouraged to similarly hold faculty and trainee sessions to review new evaluation forms, rating scales, and expectations on completion.

2) <u>Administrative Process.</u>

EFS also addressed the need for an effective mechanism to distribute and collect evaluations at the end of each resident rotation, or at established intervals for longitudinal experiences such as continuity clinics. The group determined that the New Innovations residency management system, already in use at our institutions, was suitable for this purpose. Electronic management of evaluations has many advantages, including:

- facilitating real-time review by program leadership with the "automatic notification of low ratings" New Innovations function
- allowing for trainee access to evaluations (as required by ACGME)

- tracking completion rates, timing of completion, and individual faculty compliance
- facilitating confidentiality by allowing for selective access
- allowing for analysis of summary data and trends over time for individuals, programs or institutions

During the initial implementation, where a few small fellowship programs have preferred paper-based or other on-line systems, this has been allowed.

Administrative support via GME office personnel is available to programs to support implementation of the assessment tools. For example, an expert in the New Innovations system helps program coordinators load these into the system for automatic distribution according to the trainee and attending schedules. A New Innovations "Instructor's Guide" was created and distributed to all programs to assist with implementation.

3) Faculty Development

The roll-out of new template evaluation forms creates an opportunity to re-focus faculty (and other evaluators) on the importance of evaluation and feedback, and to provide training aimed at enhancing this process. A GME Education Specialist provides specialty-specific teaching sessions focused on competency-based feedback and evaluation (based on the "Faculty Evaluation and Feedback to Trainee" evaluation) incorporating published milestones, when available. These teaching sessions are delivered as workshops, retreats, department-specific conferences, and/or grand rounds.

Faculty development sessions focused on evaluation usually occur based on departmental interest, but are also initiated when the GME office notes a problem with resident evaluation via ACGME citation, resident survey or internal review. Twenty-four sessions have been provided since 2011, with over 800 faculty and trainee participants. Evaluations of these sessions are generally excellent. Since 2013, these specialty-specific conferences and workshops have also addressed how competency-based formative evaluation informs the assessment of specialty-specific milestones. Individual faculty coaching regarding assessment is also available, but has not yet been utilized.

Limitations: Limitations of these evaluation tools and associated processes must be noted. The tools have been in use for approximately 1.5 years and have not yet been utilized or evaluated outside of the health system by which they were developed. Though information suggests a positive impact based on implementation of these tools among 36 ACGME programs and 20 non-ACGME programs at Massachusetts General Hospital and Brigham and Women's hospitals, along with the administrative processes and faculty development efforts described here, a formal outcomes analysis is not yet available.

Key Lessons learned

- Change and standardization can each be unwelcome to program directors and staff that are satisfied with their current evaluation tools. We found that abundant communication prior to implementation was important, specifically addressing: a) goals, b) expectations (including what is required by ACGME and by the institution), and c) resources available via the GME to support program implementation of the tools.

- Thorough communication and training on assessment is essential for all evaluators.

- Confidentiality

Implementation recommendations:

Implementation

The ACGME requires that trainees be assessed using multi-source evaluation, including evaluation of trainees by faculty, peers, patients, other health care professionals and self. Residency and fellowship programs should match each trainee to each evaluator based on their clinical rotation schedule and assign evaluations which should be completed at the end of each rotation. While matching trainee to evaluator is most efficiently accomplished using an on-line evaluation system (e.g. New Innovations), programs can also utilize the evaluations in a paper format. The self-assessment evaluation can be utilized at any level of training throughout residency and fellowship. It is recommended that trainees complete the self -assessment prior to their Program Director meetings and review content with their program directors or designate, as assigned by the program director.

Departments should communicate to their faculty clear expectations about how soon after the end of a block rotation – or the planned evaluation date for a longitudinal experience - evaluations should be completed. In addition, faculty should be encouraged to use the completed evaluation forms as a foundation for verbal feedback, which should also occur at the end of the rotation. (Linking completion of the evaluation tool with provision of feedback makes it increasingly important to have evaluations completed by the end of a rotation, since the resident and their evaluator(s) may no longer be in close proximity.)

Even though evaluation templates can be customized by the program, program directors and faculty should be reminded to consistently link the evaluation to the goals and objectives of the rotation—both when completing the evaluation and when providing feedback to the trainees. Connecting goals and assessment will be particularly important as milestones are implemented as part of the "Next Accreditation System".

Programs should develop explicit processes to ensure timely review, including a system for immediately conveying significant concerns to the program director.

Within 18 months of making standardized template evaluations available to GME programs in our health system, 36 of 111 ACGME programs (32 %) and 20 of 143 non-ACGME programs (14%) in our healthcare system are using these assessment tools. Implementation of forms has gone smoothly. In the context of voluntary implementation of standardized evaluations, the majority of programs have continued to use their own evaluations as were previously in use.

Acknowledgments:

Members of the Evaluation and Feedback subcommittee 2011-2013

References:

1. 2012-2013 ACGME Resident Survey, United States-Aggregated Program Data, 2013 Accreditation Council for Graduate Medical Education (ACGME).

2. ACGME Common Program Requirements 2013.

http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs2013.pdf (accessed October 1, 2013).

3. Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for Competency Assessment Programmes. Studies in Educational Evaluation, 32, 153-170.

4. Baker K. Determining resident clinical performance: getting beyond the noise. *Anesthesiology*. Oct 2011;115(4):862-878.

5. Cook, DA, Beckman, TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. (2006) Am J Med. Feb:119(2): 166.e7-166.e16.

6. Epstein RM. Assessment in medical education. *N Engl J Med*. Jan 25 2007;356(4):387-396.

7. Kogan, JR Faculty staff perceptions of feedback to residents after direct observation of clinical skills. Med Educ 2012:46(2):201-215.

8. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system--rationale and benefits. *N Engl J Med*. Mar 15 2012;366(11):1051-1056

9. Sullivan, G., Simpson, D, Cooney, T, and Beresin, E. (2013) A Milestone in the Milestones Movement: the *JGME* Milestones Supplement. Journal of Graduate Medical Education: March 2013, Vol. 5, No. 1s1, pp. 1-4.

10. Swing, SR, Beeson, MS, Carraccio, C., Coburn, M., Iobst, W., Selden, NR, Stern, PJ, Vydareny, K. (*2013*) Educational Milestone Development in the First 7 Specialties to Enter the Next Accreditation System. Journal of Graduate Medical Education: March 2013, Vol. 5, No. 1, pp. 98-106.

11. Swing SR, Clyman SG, Holmboe ES, Williams RG. Advancing resident assessment in graduate medical education. *Journal of Graduate Medical Education.* Dec 2009;1(2):278-286.

File Names of Assessment Tools:

1. Instructions for Template 360 Evaluation Forms.pdf
2. Faculty Eval of Trainee-Short version-Expectations Scale.pdf
3. Faculty Eval of Trainee-Standard version-Expectations Scale.pdf
4. Other Healthcare Professional Eval of Trainee.pdf
5. Peer Evaluation of Trainee.pdf
6. ACGME Trainee Self-Assessment.pdf
7. Faculty Evaluation of Trainee Appendix Expect Scale.pdf
8. Evaluation and Feedback Subcommittee.doc