Assessment Tools/Processes for Collecting Milestones Data

Readers' Guide

The abstracts in this document are organized by year of publication (in descending order). Readers with a particular interest in a subject may use the search function in the PDF to quickly find topics of their choosing. Please note that since the following content comes from a variety of sources, there may be variations in style and spelling. For official ACGME terminology and terminology usage, refer to the ACGME Glossary of Terms.

Background:

The implementation of the Milestones in 2009 created challenges for some training programs that needed a proper form of assessment and lacked the appropriate tools reflecting the six competencies in a reliable or valid way. Milestones created a shift in how programs evaluate residents and programs questioned how the Milestones fit into other assessment tools. Program directors and faculty were asked to make accurate milestone-based assessments but may have lacked appropriate guidance in assessment. The intention of the Milestones was not to be used as the only assessment tool. Milestones should inform the use and development of assessment tools aligned with the program's curricular goals and move residents toward unsupervised practice.

What's in the literature?

The Bibliography includes innovative methods of assessing Milestones, how to use assessment tools that are designed to provide helpful feedback to residents, and strategies for using this bibliography that may help readers in selecting assessment tools when evaluating residents. There is no single assessment tool or method used to determine readiness for unsupervised practice of residents and fellows. Residency programs can choose a combination of assessments to effectively cover the Milestones in their discipline.

Some of the assessment tools included in the literature:

• Evidence-based remediation toolkit
• Resident-based evaluation tool to facilitate earlier identification of struggling interns
• Direct Observation tools
• History of the 2001 Outcomes Project

Best Practices for Remediation in Pulmonary and Critical Care Medicine Fellowship Training

BACKGROUND:
Remediation of struggling learners in pulmonary and critical care fellowship programs is a challenge, even for experienced medical educators.

OBJECTIVE:
This evidence-based narrative review provides a framework program leaders may use to address fellows having difficulty achieving competency during fellowship training.

METHODS:
The relevant evidence for approaches on the basis of each learner's needs is reviewed and interpreted in the context of fellowship training in pulmonary medicine and critical care. Issues addressed include bias in fellow assessments and remediation, the impacts of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, the specific challenges of pulmonary and critical care fellowship programs, a brief review of relevant legal issues, guidance on building and leveraging program resources, and a discussion of learner outcomes.

RESULTS:
This results in a concise, evidence-based toolkit for program leaders based around four pillars: early identification, fellow assessment, collaborative intervention, and reassessment. Important concepts also include the need for documentation, clear and written communication, and fellow-directed approaches to the creation of achievable goals.

CONCLUSIONS:
Evidence-based remediation helps struggling learners in pulmonary and critical care fellowship to improve their ability to meet Accreditation Council for Graduate Medical Education (ACGME) milestones.

ACGME Milestones in the Real World: A Qualitative Study Exploring Response Process Evidence

BACKGROUND:
Since the Accreditation Council for Graduate Medical Education (ACGME) introduced the Milestones in 2013, the body of validity evidence supporting their use has grown, but there is a gap with regard to response process.

OBJECTIVE:
The purpose of this study is to qualitatively explore validity evidence pertaining to the response process of individual Clinical Competency Committee (CCC) members when assigning Milestone ratings to a resident.

METHODS:
Using a constructivist paradigm, we conducted a thematic analysis of semi-structured interviews with 8 Transitional Year (TY) CCC members from 4 programs immediately following a CCC meeting between November and December 2020. Participants were queried about their response process in their application of Milestone assessment. Analysis was iterative, including coding, constant comparison, and theming.

RESULTS:
Participant interviews identified an absence of formal training and a perception that Milestones are a tool for resident assessment without recognizing their role in program evaluation. In describing their thought process, participants reported comparing averaged assessment data to peers and time in training to generate Milestone ratings. Meaningful narrative comments, when available, differentiated resident performance from peers. When assessment data were absent, participants assumed an average performance.

CONCLUSIONS:
Our study found that the response process used by TY CCC members was not always consistent with the dual purpose of the Milestones to improve educational outcomes at the levels of residents and the program.

Milestones as a Faculty Development Tool for Career Academic Physicians

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) has implemented milestones for progression of residents. Career academic physicians would benefit from similar concrete guidance for scholarly activity and faculty development. After developing milestones across six recognized competencies among our family medicine academicians, we acknowledged the potential benefit of expanding the development of milestones throughout the academic medical center.

METHODS:
Milestones that we previously developed were modified by departmental leaders within our institution reflecting levels of career development based on benchmarks in each field. These objective measures for guiding maturation of clinical and academic skill sets were then circulated to clinicians in five residency programs throughout our academic medical center for self-evaluation. We analyzed the completed surveys to determine if an association exists between years in academics and rank across each area of competency.

RESULTS:
We received fifty-three responses from the 91 faculty invited. We noted a significant association in the competency of medical knowledge with progression from assistant to full professor, and we noted a trend toward significance in professionalism and progression from assistant to full professor. These objective measures of clinician development and competency suggest association with levels of academic career development by rank within the institution.

CONCLUSIONS:
This rubric can be helpful for directing faculty development and faculty mentorship. These milestones are general enough that other physician specialties may be able to adopt them for their own needs.

Natural Language Processing to Estimate Clinical Competency Committee Ratings

OBJECTIVE:
Residency program faculty participate in clinical competency committee (CCC) meetings, which are designed to evaluate residents' performance and aid in the development of individualized learning plans. In preparation for the CCC meetings, faculty members synthesize performance information from a variety of sources. Natural language processing (NLP), a form of artificial intelligence, might facilitate these complex holistic reviews. However, there is little research involving the application of this technology to resident performance assessments. With this study, we examine whether NLP can be used to estimate CCC ratings.

DESIGN:
We analyzed end-of-rotation assessments and CCC assessments for all surgical residents who trained at one institution between 2014 and 2018. We created models of end-of-rotation assessment ratings and text to predict dichotomized CCC assessment ratings for 16 Accreditation Council for Graduate Medical Education (ACGME) Milestones. We compared the performance of models with and without predictors derived from NLP of end-of-rotation assessment text.

RESULTS:
We analyzed 594 end-of-rotation assessments and 97 CCC assessments for 24 general surgery residents. The mean (standard deviation) for area under the receiver operating characteristic curve (AUC) was 0.84 (0.05) for models with only non-NLP predictors, 0.83 (0.06) for models with only NLP predictors, and 0.87 (0.05) for models with both NLP and non-NLP predictors.

CONCLUSIONS:
NLP can identify language correlated with specific ACGME Milestone ratings. In preparation for CCC meetings, faculty could use information automatically extracted from text to focus attention on residents who might benefit from additional support and guide the development of educational interventions.

Current Milestones Assessment Practices, Needs, and Challenges of Program Directors: A Collective Case Study in a Pediatric Hospital Setting

INTRODUCTION:
Accreditation Council for Graduate Medical Education's (ACGME's) Milestones assessment requirement has placed new demands on Program Directors (PDs), especially those with limited knowledge of assessment and evaluation activities. There is a lack of clarity on how Program Director (PDs)/Associate PDs (APDs) are effectively implementing milestones assessment and evaluation practices in the Graduate Medical Education programs. The purpose of this study was to investigate current assessment practices, needs, and challenges of PDs in implementing milestones assessment within their residency and fellowship programs in a pediatric hospital setting.

METHODS:
This study used a collective case study approach to obtain information from PDs, APDs, and Clinical Competency Committee (CCC) Chairs in 19 graduate programs at a pediatric hospital. We used structured meetings with planned agendas and a pre- formatted template to itemize program needs/difficulties/challenges in the milestone assessment. We used cross-case thematic content anal-ysis to identify categories and themes to compare differences and commonalities across programs.

RESULTS:
A total of 38 PDs, APDs, and CCC Chairs from 19 different specialties/subspe-cialties participated in this study. Thirteen types of assessment and evaluation tools were consistently used across programs. Three categories emerged in relation to those assessment and evaluation types (direct, indirect, and multi-source). Rotation evaluation (84.2%), direct observation (73.2%), and 360-degree assessment (68.4%) were primarily used for measuring patient care among the six core competencies. Programs' needs varied from curriculum and assessment tool development to alignment of milestones items, and to creating a sys-tematic assessment management plan. The most common challenges were difficulties related to logistics and tracking of evaluation in the survey management system (52.6%), challenges with time management (47.3%), and difficulty in determining and interpret-ing the milestones' numbers and levels (31.5%).

CONCLUSIONS:
Milestones assessment and evaluation in medical education can be a challenge, but a priority for many training programs. Our study indicated that milestones assessment and evaluation in medical education are far more com-plex than we expect. Multiple assessment methods must be utilized to evaluate all essential competencies for accurate measurement of trainees' performance abilities. Our study uncovered several issues PDs faced during the implementation of milestones assessment and needs and challenges.

Simulation-Based Assessments and Graduating Neurology Residents' Milestones: Status Epilepticus Milestones

BACKGROUND:
The American Board of Psychiatry and Neurology and the Accreditation Council for Graduate Medical Education (ACGME) developed Milestones that provide a framework for residents' assessment. However, Milestones do not provide a description for how programs should perform assessments.

OBJECTIVES:
We evaluated graduating residents' status epilepticus (SE) identification and management skills and how they correlate with ACGME Milestones reported for epilepsy and management/treatment by their program's clinical competency committee (CCC).

METHODS:
We performed a cohort study of graduating neurology residents from 3 academic medical centers in Chicago in 2018. We evaluated residents' skills identifying and managing SE using a simulation-based assessment (26-item checklist). Simulation-based assessment scores were compared to experience (number of SE cases each resident reported identifying and managing during residency), self-confidence in identifying and managing these cases, and their end of residency Milestones assigned by a CCC based on end-of-rotation evaluations.

RESULTS:
Sixteen of 21 (76%) eligible residents participated in the study. Average SE checklist score was 15.6 of 26 checklist items correct (60%, SD 12.2%). There were no significant correlations between resident checklist performance and experience or self-confidence. The average participant's level of Milestone for epilepsy and management/treatment was high at 4.3 of 5 (SD 0.4) and 4.4 of 5 (SD 0.4), respectively. There were no significant associations between checklist skills performance and level of Milestone assigned.

CONCLUSIONS:
Simulated SE skills performance of graduating neurology residents was poor. Our study suggests that end-of-rotation evaluations alone are inadequate for assigning Milestones for high-stakes clinical skills such as identification and management of SE.

Communication Skills of Grandview/Southview Medical Center General Surgery Residents

CONTEXT:
In the transition of osteopathic programs to the single-accreditation graduate medical education (GME) system, residents are required to demonstrate skill in a set of core competencies identified by the Accreditation Council of Graduate Medical Education (ACGME) prior to graduation. Included in those core competencies are interpersonal and communication skills along with professionalism.

OBJECTIVES:
To assess strengths and weaknesses of residents' interpersonal communication skills and professionalism in the Grandview/Southview Medical Center (Dayton, OH) osteopathic general surgery program using the validated Communication Assessment Tool (CAT).

METHODS:
From November 2014 to June 2018, all patients who presented for an appointment at the Cassano General Surgery Clinic were asked by a medical assistant to complete a CAT questionnaire following their encounter with a resident physician. Patients at Cassano, an outpatient office-based facility directed to the underserved local community, are seen first by an intern, then by a 4th or 5th year resident and later by an attending physician. Patients 18 years of age or older were included; patients were excluded if they were unable to understand or read English. Patient demographics were collected, including age, gender, race/ethnicity, and previous exposure to this resident physician. Each resident's name was replaced on the CAT with a number for data analysis. The resident variables collected for this study included year of training, gender, and native language.

RESULTS:
The mean response for all CAT items was 4.5 out of 5, indicating that responses to resident performance were largely positive. Patients responded to 4 of the 14 CAT items with only excellent, very good, or good responses and no fair or poor responses. Four items had only 1 fair or poor response. The remaining 6 items received more than 1 fair or poor response: "greeted me in a way that made me feel comfortable" (#1), "talked in terms I could understand" (#8), "encouraged me to ask questions" (#10), "involved me in decisions as much as I wanted" (#11), "showed care and concern" (#13), and "spent the right amount of time with me" (#14).

CONCLUSIONS:
Attending surgeons evaluate residents in multiple areas from a doctor's perspective, but there is a potential lack of correlation between that evaluation and a patient's experience, which is paramount in osteopathic medicine. Patient responses to the CAT questionnaire can be used by program directors to identify deficiencies in milestone/competency achievement and facilitate improvement both individually and programmatically for residents according to ACGME standards.

Association Between Entrustable Professional Activities and Milestones Evaluations: Real-time Assessments Correlate With Semiannual Reviews

OBJECTIVE:
Entrustable professional activities (EPAs) have been developed to refine competency-based education. The American Board of Surgery has initiated a 2-year pilot study to evaluate the impact of EPAs on the evaluation and feedback of surgical residents. The ACGME Milestones in Surgery is a semiannual competency-based evaluation program to measure resident progression through 16 professional attributes across 8 practice domains. The correlation between these 2 evaluation tools remains unclear. The purpose of this study is to evaluate this correlation through comparison of an EPA with the corresponding elements of the ACGME Milestones.

DESIGN:
From July, 2018 to October, 2019, all residents submitting EPA evaluations for gall bladder disease were evaluated for preoperative, intraoperative, and/or postoperative entrustability. The ratings were converted to a numerical rank from 0 to 4. Milestones scores from May 2019 and November 2019 were obtained for each resident, with scores ranging from 0 to 4. The gall bladder EPA incorporates the operative PC3 and MK2 and nonoperative PC1, PC2, and ICS3 components. Spearman rank correlation was conducted to evaluate the association between each resident's median EPA ranking and his/her milestones scores.

SETTING:
SUNY Upstate Medical University, Syracuse, NY, a university-based hospital.

PARTICIPANTS:
General surgery residents.

RESULTS:
Among 24 residents, 106 intraoperative EPA evaluations were. For both the May and November milestones, significant positive correlations were noted for PC3 (correlation coefficient $\rho = 0.690$, p
< 0.001; $\rho = 0.876$, p < 0.001). Similarly, for MK2, a significant positive correlation was noted ($\rho = 0.882$, p < 0.001; $\rho = 0.759$, p < 0.001). Interestingly, significant positive correlations were also identified between the 3 nonoperative milestones and the intraoperative entrustability ranking.

CONCLUSIONS:
We observed significant correlations between EPAs for cholecystectomy and associated milestones evaluation scores. These findings indicate that EPAs may provide more timely and specific feedback than existing tools and, on aggregate, may improve upon existing formative feedback practices provided through the biannual evaluation of surgical residents.

Reported Pediatrics Milestones (Mostly) Measure Program, Not Learner Performance

PURPOSE:
Semiannually, U.S. pediatrics residency programs report resident milestone levels to the Accreditation Council for Graduate Medical Education (ACGME). The Pediatrics Milestones Assessment Collaborative (PMAC, consisting of the National Board of Medical Examiners, American Board of Pediatrics, and Association of Pediatric Program Directors) developed workplace-based assessments of 2 inferences: readiness to serve as an intern with a supervisor present (D1) and readiness to care for patients with a supervisor nearby in the pediatric inpatient setting (D2). The authors compared learner and program variance in PMAC scores with ACGME milestones.

METHOD:
The authors examined sources of variance in PMAC scores and milestones between November 2015 and May 2017 of 181 interns at 8 U.S. pediatrics residency programs using random effects models with program, competency, learner, and program × competency components.

RESULTS:
Program-related milestone variance was substantial (54% D1, 68% D2), both in comparison to learner milestone variance (22% D1, 14% D2) and program variance in the PMAC scores (12% D1, 10% D2). In contrast, learner variance represented 44% (D1) or 26% (D2) of variance in PMAC scores. Within programs, PMAC scores were positively correlated with milestones for all but one competency.

CONCLUSIONS:
PMAC assessments provided scores with little program-specific variance and were more sensitive to differences in learners within programs compared with milestones. Milestones reflected greater differences by program than by learner. This may represent program-based differences in intern performance or in use of milestones as a reporting scale. Comparing individual learner milestones without adjusting for programs is problematic.

The OSPE

OBJECTIVES:
Resident assessment tends to consist of multiple-choice examinations, even in nuanced areas, such as quality assurance. Internal medicine and many other specialties use objective structured clinical examinations, or OSCEs, to evaluate residents. We adapted the OSCE for pathology, termed the Objective Structured Pathology Examination (OSPE).

METHODS:
The OSPE was used to evaluate first- and second-year residents over 2 years. The simulation included an anatomic pathology sign-out session, where the resident could be evaluated on diagnostic skills and knowledge of key information for cancer staging reports, as well as simulated frozen-section analysis, where the resident could be evaluated on communication skills with a "surgeon." The OSPE also included smaller cases with challenging quality issues, such as mismatched slides or gross description irregularities. All cases were scored based on the Pathology Milestones created by the Accreditation Council for Graduate Medical Education.

RESULTS:
Using this OSPE, we were able to demonstrate that simulated experiences can be an appropriate tool for standardized evaluation of pathology residents.

CONCLUSIONS:
Yearly evaluation using the OSPE could be used to track the progress of both individual residents and the residency program as a whole, identifying problem areas for which further educational content can be developed.

Using Multisource Feedback to Assess Resident Communication Skills: Adding a New Dimension to Milestone Data

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) requires evaluation of residents' communication skills. These evaluations should involve assessments from a variety of persons with different perspectives and opportunities to observe resident behavior. Our objectives with this study were to determine if parents, nurses, and physicians significantly differed in their ratings of residents' communication skills; to ascertain the degree of association between these evaluations and ACGME milestone data; and to elicit feedback from residents about the specificity and usefulness of this type of evaluation compared to the evaluations they were typically provided.

METHODS:
During the 2016-2017 academic year, parents of patients ready for discharge, nurses, and attending physicians completed evaluations of resident communication skills. A repeated measures multivariate analysis of variance compared communication skills scores across the 3 groups of raters. Resident ACGME milestone ratings for interpersonal and communication skills were correlated with the communication skills evaluations. Residents rated the specificity and usefulness of the 360-degree evaluations.

RESULTS:
Parents rated residents' communication skills significantly higher than nurses and physicians rated them. We found no significant difference between the nurse and physician ratings. A significant correlation was found between resident ratings by physicians and ACGME milestone data. Residents found the feedback from these evaluations to be more specific and useful in delineating their communication strengths and weaknesses than typical milestone feedback.

CONCLUSION:
Parents added a unique perspective about residents' communication and should be included in resident evaluation when feasible. Residents appreciated the specificity and usefulness of the evaluation instrument.

Best Approaches to Evaluation and Feedback in Post-Graduate Medical Education

PURPOSE OF REVIEW:
The objectives of this literature review are to appraise current approaches and assess new technologies that have been utilized for evaluation and feedback of residents, with focus on surgical trainees.

RECENT FINDINGS:
In 1999, the Accreditation Council for Graduate Medical Education introduced the Milestone system as a tool for summative evaluation. The organization allows individual program autonomy on how evaluation and feedback are performed. In the past, questionnaire evaluations and informal verbal feedback were employed. However, with the advent of technology, they have taken a different shape in the form of crowdsourcing, mobile platforms, and simulation. Limited data is available on new methods but studies show promise citing low cost and positive impact on resident education. No one "best approach" exists for evaluation and feedback. However, it is apparent that a multimodal approach that is based on the ACGME Milestones can be effective and aid in guiding programs.

Implementation and Initial Construct Validity Evidence of a Tool, myTIPreport, for Interactive Workplace Feedback on ACGME Milestones

OBJECTIVE:
To describe implementation of myTIPreport for milestone feedback and to initiate construct validity testing of myTIPreport for milestones.

DESIGN:
myTIPreport was used to provide workplace feedback on Accreditation Council for Graduate Medical Education required milestone sets. Performance of senior learners (postgraduate year [PGY]-4s) was compared to that of junior learners (PGY-1s) to begin the process of construct validity testing for myTIPreport.

SETTING:
A convenience-based site selection of Obstetrics and Gynecology (OBGYN) residency programs.

PARTICIPANTS:
OBGYN residents and faculty.

RESULTS:
Amongst the 12 participating OBGYN residency programs, there were 444 unique learners and 343 unique faculty teachers. A total of 5293 milestone feedback encounters were recorded. Mean PGY-4 performance was rated higher than mean PGY-1 performance on all 25 of the compared milestone sets, with statistically significant differences seen for 19 (76%) of these 25 milestone sets and nonsignificant differences in the predicted direction observed for the other 6 milestone sets.

CONCLUSIONS:
myTIPreport detected differences between senior and junior learners for the majority of compared feedback encounters for OBGYN residents. Findings support the emerging construct validity of myTIPreport for milestone feedback.

Improving Our Ability to Predict Resident Applicant Performance: Validity Evidence for a Situational Judgment Test

Cullen MJ, Zhang C, Marcus-Blank B, Braman JP, Tiryaki E, Konia M, Hunt MA, Lee MS, Heest AV, Englander R, Sackett PR, Andrews JS. Teach Learn Med. 2020 May 19;1-14. doi: 10.1080/10401334.2020.1760104.

CONSTRUCT:
We investigated whether a situational judgment test (SJT) designed to measure professionalism in physicians predicts residents' performance on (a) Accreditation Council for Graduate Medical Education (POACGME) competencies and (b) a multisource professionalism assessment (MPA).

BACKGROUND:
There is a consensus regarding the importance of assessing professionalism and interpersonal and communication skills in medical students, residents, and practicing physicians. Nonetheless, these noncognitive competencies are not well measured during medical education selection processes. One promising method for measuring these noncognitive competencies is the SJT. In a typical SJT, respondents are presented with written or video-based scenarios and asked to make choices from a set of alternative courses of action. Interpersonally oriented SJTs are commonly used for selection to medical schools in the United Kingdom and Belgium and for postgraduate selection of trainees to medical practice in Belgium, Singapore, Canada, and Australia. However, despite international evidence suggesting that SJTs are useful predictors of in-training performance, end-of-training performance, supervisory ratings of performance, and clinical skills licensing objective structured clinical examinations, the use of interpersonally oriented SJTs in residency settings in the United States has been infrequently investigated. The purpose of this study was to investigate whether residents' performance on an SJT designed to measure professionalism-related competencies-conscientiousness, integrity, accountability, aspiring to excellence, teamwork, stress tolerance, and patient-centered care-predicts both their current and future performance as residents on two important but conceptually distinct criteria: ACGME competencies and the MPA.

APPROACH:
We developed an SJT to measure seven dimensions of professionalism. During calendar year 2017, 21 residency programs from 2 institutions administered the SJT. We conducted analyses to determine the validity of SJT and USMLE scores in predicting milestone performance in ACGME core competency domains and the MPA in June 2017 and 3 months later in September 2017 for the MPA and 1 year later, in June 2018, for ACGME domains.

RESULTS:
At both periods, the SJT score predicted overall ACGME milestone performance (r = .13 and .17, respectively; $p < .05$) and MPA performance (r = .19 and .21, respectively; $p < .05$). In addition, the SJT predicted ACGME patient care, systems-based practice, practice-based learning and improvement, interpersonal and communication skills, and professionalism competencies (r = .16, .15, .15, .17, and .16, respectively; $p < .05$) 1 year later. The SJT score contributed incremental validity over USMLE scores in predicting overall ACGME milestone performance (ΔR = .07) 1 year later and MPA performance (ΔR = .05) 3 months later.

CONCLUSIONS:
SJTs show promise as a method for assessing noncognitive attributes in residency program applicants. The SJT's incremental validity to the USMLE series in this study underscores the importance of moving beyond these standardized tests to a more holistic review of candidates that includes both cognitive and noncognitive measures.

Linking Workplace-Based Assessment to ACGME Milestones: A Comparison of Mapping Strategies in Two Specialties

CONSTRUCT:
The construct that is assessed is competency in Pediatrics and Internal Medicine residency training. Background: The Accreditation Council for Graduate Medical Education (ACGME) created milestones to measure learner progression toward competence over time but not as direct assessment tools. Ideal measurement of resident performance includes direct observation and assessment of patient care skills in the workplace. Residency programs have linked these concepts by mapping workplace-based assessments to the milestones of ACGME subcompetencies. Mapping is a subjective process, and little is known about specific techniques or the resulting consequences of mapping program-specific assessment data to larger frameworks of competency.

APPROACH:
In this article, the authors compare and contrast the techniques used to link workplace-based assessments called Observable Practice Activities (OPAs) to ACGME subcompetencies in two large academic residency programs from different specialties (Internal Medicine and Pediatrics). Descriptive analysis explored the similarities and differences in the assessment data generated by mapping assessment items to larger frameworks of competency.

RESULTS:
Each program assessed the core competencies with similar frequencies. The largest discrepancy between the two subspecialties was the assessment of Medical Knowledge, which Internal Medicine assessed twice as often. Pediatrics also assessed the core competency Systems-based Practice almost twice as often as Internal Medicine. Both programs had several subcompetencies that were assessed more or less often than what appeared to be emphasized by the blueprint of mapping. Despite using independent mapping processes, both programs mapped each OPA to approximately three subcompetencies.

CONCLUSIONS:
Mapping workplace-based assessments to the ACGME subcompetencies allowed each program to see the whole of their curricula in ways that were not possible before and to identify existing curricular and assessment gaps. Although each program used similar assessment tools, the assessment data generated were different. The lessons learned in this work could inform other programs attempting to link their own workplace-based assessment elements to ACGME subcompetencies.

Application Factors May Not Be Predictors of Success Among General Surgery Residents as Measured by ACGME Milestones

BACKGROUND:
Can factors within the Electronic Residency Application Service application be used to predict the success of general surgery residents as measured by the Accreditation Council for Graduate Medical Education (ACGME) general surgery milestones?

METHODS:
This is a retrospective study of 21 residents who completed training at a single general surgery residency program. Electronic Residency Application Service applications were reviewed for objective data, such as age, US Medical Licensing Examination scores, and authorship of academic publications as well as for letters of recommendation, which were scored using a standardized grading system. These factors were correlated to resident success as measured by ACGME general surgery milestone outcomes using univariate and multivariate analyses. This study was conducted at a single academic tertiary care and level 1 trauma facility. Residents who completed general surgery residency training from the years of 2012-2018 were included in the study.

RESULTS:
There were few correlations between application factors and resident success determined by the ACGME milestones.

CONCLUSIONS:
Application factors alone do not account for ongoing growth and development throughout residency. Unlike the results presented in the literature for other surgical subspecialties, predicting general surgery resident success based on application factors is not straightforward.

Development and Establishment of Initial Validity Evidence for a Novel Tool for Assessing Trainee Admission Notes

BACKGROUND:
Documentation is a key component of practice, yet few curricula have been published to teach trainees proper note construction. Additionally, a gold standard for assessing note quality does not exist, and no documentation assessment tools integrate with established competency-based frameworks.

OBJECTIVE:
To develop and establish initial validity evidence for a novel tool that assesses key components of trainee admission notes and maps to the Accreditation Council for Graduate Medical Education (ACGME) milestone framework.

DESIGN:
Using an iterative, consensus building process we developed the Admission Note Assessment Tool (ANAT). Pilot testing was performed with both the supervising attending and study team raters not involved in care of the patients. The finalized tool was piloted with attendings from other institutions.

PARTICIPANTS:
Local experts participated in tool development and pilot testing. Additional attending physicians participated in pilot testing.

MAIN MEASURES:
Content, response process, and internal structure validity evidence was gathered using Messick's framework. Inter-rater reliability was assessed using percent agreement.

KEY RESULTS:
The final tool consists of 16 checklist items and two global assessment items. Pilot testing demonstrated rater agreement of 72% to 100% for checklist items and 63% to 70% for global assessment items. Note assessment required an average of 12.3 min (SD 3.7). The study generated validity evidence in the domains of content, response process, and internal structure for use of the tool in rating admission notes.

CONCLUSIONS:
The ANAT assesses individual components of a note, incorporates billing criteria, targets note "bloat," allows for narrative feedback, and provides global assessments mapped to the ACGME milestone framework. The ANAT can be used to assess admission notes by any attending and at any time after note completion with minimal rater training. The ANAT allows programs to implement routine note assessment for multiple functions with the use of a single tool.

Developing a Novel Scoring System to Objectively Track Orthopaedic Resident Educational Performance and Progression

OBJECTIVE:
Objectively determining orthopedic resident competence remains difficult and lacks standardization across residency programs. We sought to develop a scoring system to measure resident educational activity to stratify participation and performance in particular aspects of training and the effect of these measures on board certification.

DESIGN:
A weighted scoring system (Average Resident Score, ARS) was developed using the number of logged cases, clinic notes dictated, OITE PGY percentile, case minimums met, and scholarly activity completed each academic year (AY), with clinical activity being more heavily weighted. The Resident Effectiveness Score (RES), a z-score showing the number of standard deviations from the mean, was determined using the ARS. The RES effect on the Accreditation Council for Graduate Medical Education (ACGME) Milestones and American Board of Orthopedic Surgery (ABOS) Part 1 percentile score was determined using a Spearman correlation.

SETTING:
Large academic orthopedic residency.

PARTICIPANTS:
Thirty one orthopedic residents graduating between 2011 and 2016 were included.

RESULTS:
The RES did not differ between classes in the same AY, nor change significantly for individual residents during their training. Milestone z-scores increased as residents progressed in their education. The RES correlated with each Milestone competency subscore. The PGY5 OITE score and achieving ACGME minimums correlated with passing ABOS Part 1 (28/31 1st time pass), but the RES did not predict passing the board examination.

CONCLUSIONS:
This study demonstrates a scoring system encompassing multiple facets of resident education to track resident activity and progress. The RES can be tailored to an individual program's goals and aims and help program directors identify residents not maximizing educational opportunities compared to their peers. Monitoring this score may allow tailoring of educational efforts to individual resident needs. This RES may also allow residents to measure their performance and educational accomplishments and adjust their focus to obtain competence and board certification.

Advancing Competency-Based Medical Education Through Assessment and Feedback in Breast Imaging

Anna I Holbrook , Claudia Kasales . 2020 Mar;27(3):442-446. doi: 10.1016/j.acra.2019.04.017. Epub 2019 May 27.

ABSTRACT:
Competency-based medical education (CBME) is a method of educating and assessing trainees that focuses on outcomes, rather than process. In this review, we inform radiologists involved in breast imaging training on the tenets of CBME and its relationship to the milestones, feedback and assessment. We also describe multiple methods for assessment specific to the breast imaging curriculum, and techniques for improving feedback to trainees in breast imaging.

Content Analysis of Family Medicine Resident Peer Observations

BACKGROUND AND OBJECTIVES:
Direct observation is a critical part of assessing learners' achievement of the Accreditation Council for Graduate Medical Education (ACGME) Milestones and subcompetencies. Little research exists identifying the content of peer feedback among residents; this study explored the content of residents' peer assessments as they relate to ACGME Milestone subcompetencies in a family medicine residency program.

METHODS:
Using content from a mobile app-based observation tool (M3App), we examined resident peer observations recorded between June 2014 and November 2017, tabulating frequency of observation for each ACGME subcompetency and calculating the proportion of observations categorized under each subcompetency, as well as for each postgraduate year (PGY) class. We also coded each observation on three separate dimensions: "positive," "constructive," and "actionable." We used the $\chi 2$ test for independence, and estimated odds ratios and 95% confidence intervals for two-by-two comparisons to compare numbers of observations within each category.

RESULTS:
Our data include 886 peer observations made by 54 individual residents. The most frequently observed competencies were in patient care, communication, and professionalism (56%, 47%, and 38% of observations, respectively). Practice-based learning and improvement was observed least frequently (16% of observations). On average, 97.25% of the observations were positive, 85% were actionable, and 6% were constructive.

CONCLUSIONS:
When asked to review their peers, residents provide comments that are primarily positive and actionable. In addition, residents tend to provide more feedback on certain subcompetencies compared to others, suggesting that programs may rely on peer feedback for specific subcompetencies. Peers can provide perspective on the behaviors and skills of fellow residents.

Professionalism and Ethics: A Standardized Patient Observed Standardized Clinical Examination to Assess ACGME Pediatric Professionalism Milestones

INTRODUCTION:
The ethical skills fundamental to medical practice encompass a large portion of the Accreditation Council for Graduate Medical Education (ACGME) professionalism milestones. Yet many ethical practices are difficult to reduce to milestone frameworks given the variety of traditions of moral reasoning that clinician-trainees and their colleagues might properly employ.

METHODS:
We developed an observed standardized clinical examination (OSCE) simulation with standardized patients to assess the ethical skills captured in professionalism milestones in pediatrics. The OSCE included four vignettes based on actual cases that presented problems without a correct answer. Residents discussed ethically challenging issues with standardized patients and were evaluated on specific ethical tenets contained in the professionalism milestones. Our assessment guide for preceptors offered content for debriefing and assessment. We piloted this OSCE with seven preceptors and 17 pediatric residents in two different medical settings.

RESULTS:
Residents all agreed that the four cases were realistic. All but two residents agreed that OSCEs like this one are an appropriate or objective way of assessing the ACGME professionalism milestones. All preceptors reported that they strongly agreed the assessment improved their ability to assess the professionalism milestones.

DISCUSSION:
This OSCE offers a structured method to assess professionalism milestones and a forum to discuss ethical problem solving. It can also be used solely as a training exercise in ethical decision making and having difficult conversations.

Development of Subspecialty-Specific Reporting Milestones for Hospice and Palliative Medicine Fellowship Training in the US

Barnett M, Buckholz G, Christensen A, Hwang J, Johnston CB, Landzaat L, Lupu D, Morrison LJ, Okon T, Radwany S, Yang H, Edgar L, Gustin J. J Pain Symptom Manage. 2020 Jan 24. pii: S0885- 3924(20)30060-9. doi: 10.1016/j.jpainsymman.2020.01.008. [Epub ahead of print]

ABSTRACT:
Continuing the transition to competency-based education, Hospice and Palliative Medicine (HPM) fellowship programs began using context-free reporting milestones (RMs) for Internal Medicine subspecialties in 2014 but quickly recognized they did not reflect the nuanced practice of the field. This article describes the development of 20 subspecialty-specific RMs through consensus group process and vetting by HPM educators. A workgroup of content experts employed an iterative consensus building process between December 2017 and February 2019 to draft new RMs and to create a Supplemental Guide that outlines the intent of each RM, examples of each developmental trajectory, assessment methods, and resources to guide educators. Program directors, program coordinators, and designated institutional officers were contacted directly to solicit feedback. The majority of respondents agreed or strongly agreed that each RM represented a realistic progression of knowledge, skills, and behaviors, and that the set of milestones adequately discriminated between meaningful levels of competency. Similarly, respondents felt that the Supplemental Guide was a useful resource. The result is a set of carefully developed and broadly vetted RMs that represent a progression of development for HPM physicians over one year of clinical fellowship training.

Mini-Clinical Evaluation Exercise in the Era of Milestones and Entrustable Professional Activities in Obstetrics and Gynaecology: Resume or Reform?

OBJECTIVE:
The Accreditation Council for Graduate Medical Education (ACGME) milestones and the core Entrustable Professional Activities (EPAs) provide guiding frameworks and requirements for assessing residents' progress. The Mini-Clinical Evaluation Exercise (Mini-CEX) is a formative assessment tool used to provide direct observation after an ambulatory or clinical encounter. This study aimed to investigate the feasibility and reliability of the Mini-CEX in the authors' obstetrics and gynecology (OB/GYN) residency program and its ability to measure residents' progress and competencies in the frameworks of ACGME milestones and EPAs.

METHODS:
OB/GYN residents' 5-academic-year Mini-CEX performance was analyzed retrospectively to measure reliability and feasibility. Additionally, realistic evaluation was conducted to assess the usefulness of Mini-CEX in the frameworks of ACGME milestones and EPAs.

RESULTS:
A total of 395 Mini-CEX evaluations for 49 OB/GYN residents were analyzed. Mini-CEX evaluation data significantly discriminated among residents' training levels (P < 0.003). Residents had an average of 8.1 evaluations per resident completed; 10% of second-year residents and 28% of third-year residents were evaluated 10 or more times per year, whereas no post-graduate year 1 or post- graduate year 4 residents achieved this number. Mini-CEX data could contribute to all six primary measurement domains of OB/GYN milestones and eight of 10 EPAs required for first-year residents.

CONCLUSION:
The Mini-CEX demonstrated potential for measuring residents' clinical competencies in their ACGME milestones. Faculty time commitment was the main challenge. Reform is necessary for the current feedback structure in Mini-CEX, faculty development, and operational guidelines that help residency programs match residents' clinical competency ratings with ACGME milestones and EPAs.

Understanding Assessment Systems for Clinical Competency Committee Decisions: Evidence from a Multisite Study of Psychiatry Residency Training Programs

OBJECTIVE:
This multisite study examines how clinical competency committees in Psychiatry synthesize resident assessments to inform milestones decisions to provide guidelines that support their use.

METHODS:
The study convened training directors and associate training directors from three psychiatry residency programs to examine decision-making processes of clinical competency committees. Annual resident assessments for one second year and one third year resident were used in a mock clinical competency committee format to assign milestones for two consecutive reporting periods. The committees reflected on the process and rated how the assessment tools impacted the assessment of milestones and evaluated the overall process. The authors compared reliability of assessment between the mock committees and examined both reliability of end of rotation assessments and their composite scores when combined with clinical skills evaluations.

RESULTS:
End of rotation evaluations were the most informative tool for assigning milestones and clarifying discrepancies in performance. In particular, the patient care and medical knowledge competencies were the easiest to rate, while the systems-based practice and practice-based learning and improvement were the most difficult. Reliability between committees was low although higher number of available evaluations improved reliability in decision-making.

CONCLUSIONS:
The results indicate that the medical knowledge and patient care competencies are the easiest to rate and informed most by end of rotation evaluations and clinical skills examinations. Other evaluation tools may better capture performance on specific sub-competencies beyond workplace-based assessment, or it may be helpful to reconsider the utility of how individual sub-competencies are evaluated.

Which Emergency Medicine Milestone Sub-competencies are Identified Through Narrative Assessments?

INTRODUCTION:
Evaluators use assessment data to make judgments on resident performance within the Accreditation Council for Graduate Medical Education (ACGME) milestones framework. While workplace-based narrative assessments (WBNA) offer advantages to rating scales, validity evidence for their use in assessing the milestone sub-competencies is lacking. This study aimed to determine the frequency of sub-competencies assessed through WBNAs in an emergency medicine (EM) residency program.

METHODS:
We performed a retrospective analysis of WBNAs of postgraduate year (PGY) 2-4 residents. A shared mental model was established by reading and discussing the milestones framework, and we created a guide for coding WBNAs to the milestone sub-competencies in an iterative process. Once inter-rater reliability was satisfactory, raters coded each WBNA to the 23 EM milestone sub-competencies.

RESULTS:
We analyzed 2517 WBNAs. An average of 2.04 sub-competencies were assessed per WBNA. The sub-competencies most frequently identified were multitasking, medical knowledge, practice-based performance improvement, patient-centered communication, and team management. The sub- competencies least frequently identified were pharmacotherapy, airway management, anesthesia and acute pain management, goal-directed focused ultrasound, wound management, and vascular access. Overall, the frequency with which WBNAs assessed individual sub-competencies was low, with 14 of the 23 sub-competencies being assessed in less than 5% of WBNAs.

CONCLUSION:
WBNAs identify few milestone sub-competencies. Faculty assessed similar sub-competencies related to interpersonal and communication skills, practice-based learning and improvement, and medical knowledge, while neglecting sub-competencies related to patient care and procedural skills. These findings can help shape faculty development programs designed to improve assessments of specific workplace behaviors and provide more robust data for the summative assessment of residents.

A Systematic Review of the Use of Google Glass in Graduate Medical Education

BACKGROUND:
Graduate medical education (GME) has emphasized the assessment of trainee competencies and milestones; however, sufficient in-person assessment is often constrained. Using mobile hands-free devices, such as Google Glass (GG) for telemedicine, allows for remote supervision, education, and assessment of residents.

OBJECTIVE:
We reviewed available literature on the use of GG in GME in the clinical learning environment, its use for resident supervision and education, and its clinical utility and technical limitations.

METHODS:
We conducted a systematic review in accordance with 2009 PRISMA guidelines. Applicable studies were identified through a review of PubMed, MEDLINE, and Web of Science databases for articles published from January 2013 to August 2018. Two reviewers independently screened titles, abstracts, and full-text articles that reported using GG in GME and assessed the quality of the studies. A systematic review of these studies appraised the literature for descriptions of its utility in GME.

RESULTS:
Following our search and review process, 37 studies were included. The majority evaluated GG in surgical specialties (n = 23) for the purpose of surgical/procedural skills training or supervision. GG was predominantly used for video teleconferencing, and photo and video capture. Highlighted positive aspects of GG use included point-of-view broadcasting and capacity for 2-way communication. Most studies cited drawbacks that included suboptimal battery life and HIPAA concerns.

CONCLUSIONS:
GG shows some promise as a device capable of enhancing GME. Studies evaluating GG in GME are limited by small sample sizes and few quantitative data. Overall experience with use of GG in GME is generally positive.

Could Blockchain Technology Empower Patients, Improve Education, and Boost Research in Radiology Departments? An Open Question for Future Applications

ABSTRACT

Blockchain can be considered as a digital database of cryptographically validated transactions stored as blocks of data. Copies of the database are distributed on a peer-to-peer network adhering to a consensus protocol for authentication of new blocks into the chain. While confined to financial applications in the past, this technology is quickly becoming a hot topic in healthcare and scientific research. Potential applications in radiology range from upgraded monitoring of training milestones achievement for residents to improved control of clinical imaging data and easier creation of secure shared databases.

Validation of Assessing Arthroscopic Skill using the ASSET Evaluation

BACKGROUND:

The Accreditation Council for Graduate Medical Education and the American Board of Orthopaedic Surgery have implemented "milestones" to evaluate residents during their progression in medical education. The purpose of this study was to determine whether a validated evaluation tool correlates with surgical experience, year in training, and progression over time.

DESIGN:

This was a retrospective study of already collected curriculum assessment data where 2 unbiased, blinded orthopedic surgeons evaluated resident performance on basic diagnostic knee arthroscopy using the Arthroscopic Surgical Skills Evaluation Tool (ASSET) over 3 years. Residents also gained arthroscopy experience through a structured arthroscopy curriculum and clinical experience.

SETTING:

The study was conducted at the TRIA Orthopaedic Center (Bloomington, Minnesota, USA), an institutional site for The University of Minnesota orthopedic surgery residency program.

PARTICIPANTS:

Eleven orthopedic surgery residents at postgraduate years 2 to 5 were evaluated using the ASSET.

RESULTS:

The Pearson's Correlation Coefficient was used to validate both the number of arthroscopic procedures performed by residents (r = 0.946) and their level in training (r = 0.89). Residents who were re-evaluated after undergoing the arthroscopy curriculum throughout the year displayed significant increases in total ASSET scores (p < 0.01).

CONCLUSION:

Resident performance on the ASSET correlated with arthroscopic experience based on year-in training. More importantly, performance improved with additional years of training, demonstrating validity over time. The data also demonstrates interobserver reliability. Due to these correlations between exposure to surgery and score on the ASSET, we believe the tool could serve as a suitable means for assessing residents' technical proficiency as required by The Accreditation Council for Graduate Medical Education program guidelines.

Using an Entrustable Professional Activity to Assess Consultation Requests Called on an Internal Medicine Teaching Service

Kang AJ, Gielissen K, Windish D. MedEdPORTAL. 2019 Nov 22;15:10854. doi: 10.15766/mep_2374- 8265.10854.

INTRODUCTION:
The Accreditation Council for Graduate Medical Education's milestones require internal medicine residents to have competency in calling consults. Based on a literature review, we developed an Entrustable Professional Activity (EPA) to delineate the knowledge, skills, and attitudes required for a consultation request and, building on the EPA, implemented an assessment instrument to provide feedback to interns calling consultation requests and assess the quality of their consult questions and the level of supervision required in performing this milestone.

METHODS:
Assessments were done on internal medicine inpatient teaching services. Consultation requests were performed by interns and observed by residents using the assessment instrument. Feedback was provided to the interns. Interns then completed a self-reflection instrument based on the feedback.

RESULTS:
Twenty-six paired observations were collected over three 1-month rotations. There was a moderate positive correlation (r = .43) comparing resident and intern responses to how they felt about the intern's ability to make a consultation request. There was a strong positive correlation (r = .65) comparing resident opinion of how strong the intern's ability in calling a consult to how well the consult question used the PICO (patient, intervention, comparators, outcomes of interest) framework. Twenty-five out of 28 interns (89%) said they would make a change during their next consultation request due to feedback from their resident.

DISCUSSION:
Our EPA-based assessment instrument provided an opportunity to give interns feedback and to assess the quality of the consultation requests they made.

Focused Teaching Improves Medical Student Professionalism and Data Gathering Skills in the Emergency Department

INTRODUCTION:
Leaders in medical education have developed milestones and core competencies in an attempt to ensure that relational skills, such as communication and professionalism, are emphasized in addition to the usual skills of medical knowledge, data gathering, and emergency stabilization during students' emergency medicine (EM) medical education. Providers facile in each of these areas have better patient outcomes, patient experiences, and decreased incidence of malpractice cases. The authors attempted to demonstrate that by deliberate teaching of these skills during an EM medical student clerkship, students could significantly improve their clinical performance.

METHODS:
This prospective, randomized, single-blinded cohort study was performed at an academic, tertiary, urban ED to investigate the effects of a one-on-one preceptor shift on the clinical performance of fourth-year medical students. Students were randomized into two groups and assessed by pre- and post-intervention objective structured clinical encounters (OSCEs) with standardized patients (SPs) at weeks one and three. A crossover design was employed so that students in the control group participated in a preceptor shift after their second OSCE. Measurements were based on a five-point Likert scale assessment linked to early EM milestones as defined by the Accreditation Council on Graduate Medical Education (ACGME).

RESULTS:
The mean improvement in total overall score was significantly greater in the intervention group: 4.31 versus 2.57 (Cohen's d = 0.57, p = 0.029). When each milestone was assessed individually, students in the intervention group improved significantly in data gathering (Cohen's d = 0.47, p = 0.048) and professionalism (Cohen's d = 0.66, p = 0.011). There was a nonstatistically significant improvement for the intervention compared to control group in emergency management and communication skills. There was no improvement for either group in medical knowledge.

CONCLUSION:
A one-on-one preceptor shift can result in a statistically significant improvement in data gathering and professionalism skills as measured by OSCEs.

'Intern Check-in Tool' to Improve Early Identification of Struggling Interns and Facilitate Feedback

BACKGROUND:
Our current system for evaluation relies on faculty and peer evaluation of intern performance relating to general ACGME milestones. However, a program may have insufficient data to accurately identify a struggling intern until several months into the academic year.

OBJECTIVE;
To develop a brief, objective, resident-based evaluation tool to facilitate earlier identification of struggling interns in pediatric and internal medicine programs.

METHODS:
The intern check-in tool (ICT) consists of 18 items with a variety of observable key skills expected for interns (refer to attached form). It is scored on a 22-point scale of objective behaviors. Chief residents meet half way through each rotation and review the tool with senior residents supervising each intern.

RESULTS:
We implemented the use of the ICT at the beginning of the academic year in July 2018. Mid-year data are still being analyzed. In January 2019, after completion of the clinical competency committee (CCC) meetings, we will perform statistical analysis to measure correlations between the ICT scores and the overall intern performance as assessed by the CCC. We will also calculate the sensitivity and specificity for a range of ICT scores and measure correlations between the ICT scores and the demographic data, including medical school quartile and USMLE scores, for each intern. The ICT allowed us to identify a struggling intern early on who had multiple high scores. Following focused feedback and mentoring from senior residents, the intern's performance improved significantly prior to the CCC. The use of the tool has also identified interns struggling with wellness, who we have been able to integrate into a counseling program.

CONCLUSIONS:
The pilot of the ICT may supplement more general milestone evaluations. Our data are preliminary but promising. This tool might be of benefit to other programs aiming to assess intern performance. It provides a forum for residents to learn the art of giving feedback and seek strategies to help their interns improve in real time.

Synthesizing and Reporting Milestones-Based Learner Analytics: Validity Evidence From a Longitudinal Cohort of Internal Medicine Residents

PURPOSE:
Coordinating and operationalizing assessment systems that effectively streamline and measure fine- grained progression of residents at various stages of graduate medical training can be challenging. This article describes development, administration, and psychometric analyses of a learner analytics system to resolve challenges in implementation of milestones by introducing the Scoring Grid Model, operationalized in an internal medicine (IM) residency program.

METHOD:
A three-year longitudinal cohort of 34 residents at the University of Illinois at Chicago College of Medicine began using this learner analytics system, from entry (July 2013) to graduation (June 2016). Scores from 23 assessments used throughout the 3-year training were synthesized using the Scoring Grid Model learner analytics system, to generate scores corresponding to the 22 reportable IM subcompetencies. A consensus model was used to develop and pilot test the model using feedback from IM faculty members and residents. Scores from the scoring grid were used to inform promotion decisions and reporting of milestone levels. Descriptive statistics and mixed-effects regression were used to examine data trends and gather validity evidence.

RESULTS:
Initial validity evidence for content, internal structure, and relations to other variables that systematically integrate assessment scores aligned with the reportable milestones framework are presented, including composite score reliability of scores generated from the learner analytics system. The scoring grid provided fine-grained learner profiles and showed predictive utility in identifying low-performing residents.

CONCLUSIONS:
The Scoring Grid Model and associated learner analytics data platform may provide a practical, reasonable solution for generating fine-grained, milestones-based profiles supporting resident progress.

Ability of Ophthalmology Residents to Self-Assess Their Performance through Established Milestones

OBJECTIVES:
Accurate self-assessment is an important aspect of practice-based learning and improvement and a critical skill for resident growth. The Accreditation Council for Graduate Medical Education mandates semiannual milestones assessments by a clinical competency committee (CCC) for all ophthalmology residents. There are six core competencies: patient care (PC), medical knowledge, systems-based practice, practice-based learning and improvement, professionalism, and interpersonal communication skills. These competencies are assessed by the milestones rubric, which has detailed behavioral anchors and are also used for trainee self-assessments. This study compares resident self-assessed (SA) and faculty CCC milestones scores.

DESIGN:
Residents completed milestones self-assessments prior to receiving individual score reports from the CCC. Correlation coefficients were calculated comparing the SA and CCC scores. In addition, statistical models were used to determine predictors of disparities and differences between the SA and CCC scores.

SETTING:
Wilmer Eye Institute, Johns Hopkins Hospital.

PARTICIPANTS:
Twenty-one residents in the Wilmer Ophthalmology Residency program from July 2014 to June 2016.

RESULTS:
Fifty-seven self-assessments were available for the analysis. For each resident's first assessment, SA and CCC scores were strongly correlated ($r \geq 0.6$ and $p < 0.05$) for four milestones, and not correlated for the remaining 20 milestones. In multivariable models, the SA and CCC scores are less disparate for medical knowledge and systems-based practice competencies compared to practice-based learning and improvement. Higher year of training, PC and professionalism competencies were predictive of statistically significant resident overestimation of scores relative to the CCC. In addition, higher CCC scores predicted statistically significant lower SA-CCC disparities and differences. SA-CCC differences did not lower to a significant extent with repeated assessments or modification to the end-of-rotation evaluation forms.

CONCLUSIONS:
Self-assessments by ophthalmology residents are not well-correlated with faculty assessments, emphasizing the need for improved and frequent timely feedback. Residents have the greatest difficulty self-assessing their professionalism and PC competency. In general, senior residents and underperforming residents have more inaccurate self-assessments.

Should Objective Structured Clinical Examinations Assist the Clinical Competency Committee in Assigning Anesthesiology Milestones Competency?

BACKGROUND:
With the integration of Objective Structured Clinical Examinations into the Anesthesiology primary board certification process, residency programs may choose to implement Objective Structured Clinical Examinations for resident skill assessment. The aim of this study was to evaluate Objective Structured Clinical Examination-based milestone assessment and compare with Clinical Competency Committee milestone assessment that is based purely on clinical evaluations.

METHODS:
An annual Objective Structured Clinical Examination event was used to obtain milestone assessment of clinical anesthesia year 0-clinical anesthesia year 3 residents for selected milestones in patient care, professionalism, and interpersonal/communication skills. The Objective Structured Clinical Examination scenarios were different for each training level. The Clinical Competency Committee evaluated each resident semiannually based on clinical evaluations of resident performance. The Clinical Competency Committee milestone assessments from 2014 to 2016 that were recorded closest to the Objective Structured Clinical Examination event (±3 months) were compared to the Objective Structured Clinical Examination milestone assessments. A total of 35 residents were included in this analysis in 3 different training cohorts: A (graduates 2016, n = 12); B (graduates 2017, n = 10); and C (graduates 2018, n = 13). All residents participated in Objective Structured Clinical Examinations because their clinical anesthesia year 0 year and Clinical Competency Committee milestone data had been reported since December 2014.

RESULTS:
Both assessment techniques indicated a competency growth proportional to the length in training. Despite limited cumulative statistics in this study, average trends in the Objective Structured Clinical Examination-Clinical Competency Committee relationship indicated: (1) a good proportionality in reflecting competency growth; (2) a grade enhancement associated with Clinical Competency Committee assessment, dominated by evaluations of junior residents (clinical anesthesia year 0- clinical anesthesia year 1); and (3) an expectation bias in Clinical Competency Committee assessment, dominated by evaluation of senior residents (clinical anesthesia year 2-clinical anesthesia year 3).

CONCLUSIONS:
Our analysis confirms the compatibility of the 2 evaluation methods in reflecting longitudinal growth. The deviation of Objective Structured Clinical Examination assessments versus Clinical Competency Committee assessments suggests that Objective Structured Clinical Examinations may be providing additional or different information on resident performance. Educators might consider using both assessment methods to provide the most reliable and valid competency assessments during residency.

An Emergency Medicine Milestone-Based Simulation Curriculum: Acute Ischemic Stroke

INTRODUCTION:
The emergency medicine (EM) resident's ability to make independent decisions in the setting of acute ischemic stroke has been reduced as a result of the involvement of multidisciplinary teams. This simulation was created to give EM residents the opportunity to independently manage the early stages of ischemic stroke and its complications.

METHODS:
A solo learner was presented with a 55-year-old male with complaints consistent with an acute stroke. The resident had to calculate stroke severity; coordinate hospital resources; discuss risks, benefits, and alternatives to thrombolysis; and deal with subsequent complications. The learner had to keep a broad differential for sudden change in mental status and consider alternative interventions. Strategies to decrease intracranial pressure needed to be implemented while obtaining neurosurgical consultation. Debriefing included discussion of expected actions in the context of the Accreditation Council for Graduate Medical Education (ACGME) milestones. Residents' review of their video performance added additional self-reflection.

RESULTS:
A total of 69 PGY 3 EM residents independently participated in this simulation over a 5-year period. Thirty-two completed a postsimulation evaluation. Nearly all learners felt that this case reflected an actual patient encounter and increased their confidence in managing stroke. The milestone-based feedback tool was completed with all learners. Anticipated actions linked to Level 1 and 2 milestones were regularly achieved while acquisition of Level 3 and 4 actions varied.

DISCUSSION:
Case actions were uniquely characterized by the ACGME milestones, which helped to delineate learners' knowledge gaps and provided concrete areas for improvement.

Development of a Simulation-Based Interprofessional Teamwork Assessment Tool

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) Milestone projects required each specialty to identify essential skills and develop means of assessment with supporting validity evidence for trainees. Several specialties rate trainees on a milestone subcompetency related to working in interprofessional teams. A tool to assess trainee competence in any role on an interprofessional team in a variety of scenarios would be valuable and suitable for simulation-based assessment.

OBJECTIVE:
We developed a tool for simulation settings that assesses interprofessional teamwork in trainees.

METHODS:
In 2015, existing tools that assess teamwork or interprofessionalism using direct observation were systematically reviewed for appropriateness, generalizability, adaptability, ease of use, and resources required. Items from these tools were included in a Delphi method with multidisciplinary pediatrics experts using an iterative process from June 2016 to January 2017 to develop an assessment tool.

RESULTS:
Thirty-one unique tools were identified. A 2-stage review narrowed this list to 5 tools, and 81 items were extracted. Twenty-two pediatrics experts participated in 4 rounds of Delphi surveys, with response rates ranging from 82% to 100%. Sixteen items reached consensus for inclusion in the final tool. A global 4-point rating scale from novice to proficient was developed.

CONCLUSIONS:
A novel tool to assess interprofessional teamwork for individual trainees in a simulated setting was developed using a systematic review and Delphi methodology. This is the first step to establish the validity evidence necessary to use this tool for competency-based assessment

Predicting Performance of First-Year Residents: Correlations between Structured Interview, Licensure Exam, and Competency Scores in a Multi-Institutional Study

Marcus-Blank B, Dahlke JA, Braman JP, Borman-Shoap E, Tiryaki E, Chipman J, Andrews JS, Sackett PR, Cullen MJ. Acad Med. 2019 Mar;94(3):378-387. doi: 10.1097/ACM.0000000000002429.

PURPOSE:
To determine whether scores on structured interview (SI) questions designed to measure noncognitive competencies in physicians (1) predict subsequent first-year resident performance on Accreditation Council for Graduate Medical Education (ACGME) milestones and (2) add incremental validity over United States Medical Licensing Examination (USMLE) Step 1 and Step 2 Clinical Knowledge (CK) scores in predicting performance.

METHOD:
The authors developed 18 behavioral description questions to measure key noncognitive competencies (e.g., teamwork). In 2013-2015, 14 programs (13 residency, 1 fellowship) from 6 institutions used subsets of these questions in their selection processes. The authors conducted analyses to determine the validity of SIs and USMLE scores in predicting first-year resident milestone performance in the ACGME's core competency domains and overall.

RESULTS:
SI scores predicted mid-year and year-end overall performance ($r = .18$ and $.19$, respectively, P
$< .05$) and year-end performance on patient care, interpersonal and communication skills, and professionalism competencies ($r = .23$, $r = .22$, and $r = .20$, respectively, $p < .05$). SI scores contributed incremental validity over USMLE scores in predicting year-end performance on patient care ($\Delta R = .05$), interpersonal and communication skills ($\Delta R = .09$), and professionalism ($\Delta R = .09$; all $P < .05$). USMLE scores contributed incremental validity over SI scores in predicting year-end performance overall and on patient care and medical knowledge.

CONCLUSIONS:
SI scores predict first-year resident year-end performance in the interpersonal and communication skills, patient care, and professionalism competency domains. Future research should investigate whether SIs predict a range of clinically relevant outcomes.

Direct Observation Assessment of Ultrasound Competency Using a Mobile Standardized Direct Observation Tool Application With Comparison to Asynchronous Quality Assurance Evaluation

OBJECTIVES:
Competency assessment is a key component of point-of-care ultrasound (POCUS) training. The purpose of this study was to design a smartphone-based standardized direct observation tool (SDOT) and to compare a faculty-observed competency assessment at the bedside with a blinded reference standard assessment in the quality assurance (QA) review of ultrasound images.

METHODS:
In this prospective, observational study, an SDOT was created using SurveyMonkey containing specific scoring and evaluation items based on the Council of Emergency Medicine Residency-Academy of Emergency Ultrasound: Consensus Document for the Emergency Ultrasound Milestone Project. Ultrasound faculty used the mobile phone-based data collection tool as an SDOT at the bedside when students, residents, and fellows were performing one of eight core POCUS examinations. Data recorded included demographic data, examination-specific data, and overall quality measures (on a scale of 1-5, with 3 and above being defined as adequate for clinical decision making), as well as interpretation and clinical knowledge. The POCUS examination itself was recorded and uploaded to QPath, a HIPAA-compliant ultrasound archive. Each examination was later reviewed by another faculty blinded to the result of the bedside evaluation. The agreement of examinations scored adequate (3 and above) in the two evaluation methods was the primary outcome.

RESULTS:
A total of 163 direct observation evaluations were collected from 23 EM residents (93 SDOTs [57%]), 14 students (51 SDOTs [31%]), and four fellows (19 SDOTs [12%]). The trainees were evaluated on completing cardiac (54 [33%]), focused assessment with sonography for trauma (34 [21%]), biliary (25 [15%]), aorta (18 [11%]), renal (12 [7%]), pelvis (eight [5%]), deep vein thrombosis (seven [4%]), and lung scan (5 [3%]). Overall, the number of observed agreements between bedside and QA assessments was 81 (87.1% of the observations) for evaluating the quality of images (scores 1 and 2 vs. scores 3, 4, and 5). The strength of agreement is considered to be "fair" ($\kappa = 0.251$ and 95% confidence interval [CI] = 0.02-0.48). Further agreement assessment demonstrated a fair agreement for images taken by residents and students and a "perfect" agreement in images taken by fellows. Overall, a "moderate" inter-rater agreement was found in 79.1% for the accuracy of interpretation of POCUS scan (e.g., true positive, false negative) during QA and bedside evaluation ($\kappa = 0.48$, 95% CI = 0.34- 0.63). Faculty at the bedside and QA assessment reached a moderate agreement on interpretations noted by residents and students and a "good" agreement on fellows' scans.

CONCLUSION:
Using a bedside SDOT through a mobile SurveyMonkey platform facilitates assessment of competency in emergency ultrasound learners and correlates well with traditional competency evaluation by asynchronous weekly image review QA.

Evaluation of a Modified Objective Structured Assessment of Technical Skills Tool for the Assessment of Pediatric Laceration Repair Performance

INTRODUCTION:
The Accreditation Council for Graduate Medical Education (ACGME) has developed milestones including procedural skills under the core competency of patient care. Progress in training is expected to be monitored by residency programs. To our knowledge, there exists no tool to evaluate pediatric resident laceration repair performance.

METHODS:
The Objective Structured Assessment of Technical Skills was adapted to evaluate resident laceration repair performance using two components: a global rating scale (GRS) and a checklist. Pediatric and family medicine residents at a tertiary care children's hospital were filmed performing a simulated laceration repair. Videos were evaluated by at least five physicians trained in laceration repair.

Concordance correlation coefficients (CCC) were calculated for the GRS and checklist scores. Scores for each resident were compared across levels of training and procedural experience. Spearman's rank order correlations were calculated to compare the checklist and GRS.

RESULTS:
Thirty residents were filmed performing laceration repair procedures. The CCC showed fair concordance across reviewers for the checklist (0.55, 95% CI: 0.38-0.69) and the GRS (0.53, 95%CI: 0.36-0.67). There was no significant difference in scores by self-reported experience or training level. There was correlation between the median GRS and checklist scores (Spearman ρ = 0.730, p < .001).

CONCLUSION:
A novel tool to evaluate resident laceration repair performance in a pediatric emergency department showed fair agreement across reviewers. The study tool is not precise enough for summative evaluation; however, it can be used to distinguish between trainees who have and have not attained competence in laceration repair for formative feedback.

Psychotherapy Competency Milestones: an Exploratory Pilot of CBT and Psychodynamic Psychotherapy Skills Acquisition in Junior Psychiatry Residents

Ravitz P, Lawson A, Fefergrad M, Rawkins S, Lancee W, Maunder R, Leszcz M, Kivlighan DM Jr. Acad Psychiatry. 2019 Feb;43(1):61-66. doi: 10.1007/s40596-018-0940-4. Epub 2018 Jun 1.

OBJECTIVE:
Psychiatry residents train in Psychodynamic Psychotherapy and Cognitive Behavioral Therapy (CBT), evidence-supported treatments used in mental health care that can facilitate clinical reasoning, foster therapeutic alliances, and improve clinical outcomes. However, empirically derived milestones are needed to evaluate competency. This exploratory pilot examined changes over 1 year of training in junior psychiatry residents' competency milestone elements in Psychodynamic Psychotherapy and CBT.

METHODS:
Seventy-nine randomly selected audio-recorded sessions from differing phases of Psychodynamic Psychotherapy and CBT with five junior residents and ten patients were rated using the Psychotherapy Process Q-sort (PQS).

RESULTS:
In both treatments, patient engagement with attention to in-session emotions improved. In CBT, residents were directive, supported patients' self-efficacy, emphasized patients' accepting responsibility for their problems, discussed homework such as thought records, and focused on termination in the concluding sessions. In Psychodynamic Psychotherapy, residents attended to emotional arousal and linked patients' feelings or perceptions to past situations or behavior. Growth and hierarchical linear modeling differentiated these treatments, with CBT v. Psychodynamic adherence to PQS modality- specific ideal elements being 52% v.19%.

CONCLUSION:
Teaching and observation using empirically derived observable psychotherapy practice behaviors is feasible and can be used to assess milestone elements for competency-based education of psychiatry trainees.

Milestones in Plastic Surgery: Attending Assessment versus Resident Assessment

BACKGROUND:
The Plastic Surgery Milestones Project was jointly conceived by the ACGME and American Board of Plastic Surgery as a tool to improve granularity in resident feedback. Resident self- evaluations were compared to attending clinical competency committee evaluations to gauge resident self-perceptions and understanding of the milestones framework.

METHODS:
Semi-annual evaluations from June 2014-2017 were analyzed and compared with corresponding resident self-evaluations from the 2015-2017 academic year at the University of Pittsburgh Medical Center. Evaluations were analyzed for overall trends in performance. The Wilcoxon Rank-Sign test was used to identify any systematic differences between evaluations. Subgroup analysis using the chi- square test was performed to determine factors that may contribute to major assessment disparity (≥1).

RESULTS:
6,207 milestones across 187 faculty evaluations and 3,139 milestones across 106 resident self- evaluations were available for review. With the exception of PGY-2 residents, residents rated themselves at a significantly lower level in the competencies of medical education and patient care. Post-graduate year, academic year timing, and ACGME competency were associated with major assessment discrepancies.

CONCLUSIONS:
Overall, resident and faculty evaluations at our program were concordant which demonstrates that residents are capable of accurately assessing their own abilities and understanding the milestones framework. Areas of discordance between resident and faculty evaluations fostered discussion between residents and faculty and have led to multiple changes in our program. The introduction of self-evaluation tools at other programs may provide them with similar benefits.

Integration of Entrustable Professional Activities with the Milestones for Emergency Medicine Residents

INTRODUCTION:
Medical education is moving toward a competency-based framework with a focus on assessment using the Accreditation Council for Graduate Medical Education Milestones. Assessment of individual competencies through milestones can be challenging. While competencies describe characteristics of the person, the entrustable professional activities (EPAs) concept refers to work-related activities.EPAs would not replace the milestones but would be linked to them, integrating these frameworks. Many core specialties have already defined EPAs for resident trainees, but EPAs have not yet been created for emergency medicine (EM). This paper describes the development of milestone-linked EPAs for EM.

METHODS:
Ten EM educators from across North America formed a consensus working group to draft EM EPAs, using a modified Glaser state-of-the-art approach. A reactor panel with EPA experts from the United States, Canada and the Netherlands was created, and an iterative process with multiple revisions was performed based on reactor panel input. Following this, the EPAs were sent to the Council of Residency Directors for EM (CORD-EM) listserv for additional feedback.

RESULTS:
The product was 11 core EPAs that every trainee from every EM program should be able to perform independently by the time of graduation. Each EPA has associated knowledge, skills, attitudes and behaviors (KSAB), which are either milestones themselves or KSABs linked to individual milestones. We recognize that individual programs may have additional focus areas or work-based activities they want their trainees to achieve by graduation; therefore, programs are also encouraged to create additional program-specific EPAs.

CONCLUSION:
This set of 11 core, EM-resident EPAs can be used as an assessment tool by EM residency programs, allowing supervising physicians to document the multiple entrustment decisions they are already making during clinical shifts with trainees. The KSAB list within each EPA could assist supervisors in giving specific, actionable feedback to trainees and allow trainees to use this list as an assessment-for- learning tool. Linking each KSAB to individual EM milestones allows EPAs to directly inform milestone assessment for clinical competency committees. These EPAs serve as another option for workplace- based assessment, and are linked to the milestones to create an integrated framework.

Assessment of Emergency Medicine Resident Performance in an Adult Simulation Using a Multisource Feedback Approach

INTRODUCTION:
The Accreditation Council for Graduate Medical Education (ACGME) specifically notes multisource feedback (MSF) as a recommended means of resident assessment in the emergency medicine (EM) Milestones. High-fidelity simulation is an environment wherein residents can receive MSF from various types of healthcare professionals. Previously, the Queen's Simulation Assessment Tool (QSAT) has been validated for faculty to assess residents in five categories: assessment; diagnostic actions; therapeutic actions; interpersonal communication, and overall assessment. We sought to determine whether the QSAT could be used to provide MSF using a standardized simulation case.

METHODS:
Prospectively after institutional review board approval, residents from a dual ACGME/ osteopathic- approved postgraduate years (PGY) 1-4 EM residency were consented for participation. We developed a standardized resuscitation after overdose case with specific 1-5 Likert anchors used by the QSAT. A PGY 2-4 resident participated in the role of team leader, who completed a QSAT as self-assessment. The team consisted of a PGY-1 peer, an emergency medical services (EMS) provider, and a nurse. Two core faculty were present to administer the simulation case and assess. Demographics were gathered from all participants completing QSATs. We analyzed QSATs by each category and on cumulative score. Hypothesis testing was performed using intra-class correlation coefficients (ICC), with 95% confidence intervals. Interpretation of ICC results was based on previously published definitions.

RESULTS:
We enrolled 34 team leader residents along with 34 nurses. A single PGY-1, a single EMS provider and two faculty were also enrolled. Faculty provided higher cumulative QSAT scores than the other sources of MSF. QSAT scores did not increase with team leader PGY level. ICC for inter-rater reliability for all sources of MSF was 0.754 (0.572-0.867). Removing the self- evaluation scores increased inter-rater reliability to 0.838 (0.733-0.910). There was lesser agreement between faculty and nurse evaluations than from the EMS or peer evaluation. Conclusion: In this single-site cohort using an internally developed simulation case, the QSAT provided MSF with excellent reliability. Self-assessment decreases the reliability of the MSF, and our data suggest self-assessment should not be a component of MSF. Use of the QSAT for MSF may be considered as a source of data for clinical competency committees.

Resident Case Volume Correlates with Clinical Performance: Finding the Sweet Spot

RATIONALE AND OBJECTIVES:
To determine whether the total number of studies interpreted during radiology residency correlates with clinical performance as measured by objective criteria.

MATERIALS AND METHODS:
We performed a retrospective cohort study of three graduating classes of radiology residents from a single residency program between the years 2015-2017. The total number of studies interpreted by each resident during residency was tracked. Clinical performance was determined by tracking an individual resident's major discordance rate. A major discordance was recorded when there was a difference between the preliminary resident interpretation and final attending interpretation that could immediately impact patient care. Accreditation council for graduate medical education milestones at the completion of residency, Diagnostic radiology in- training scores in the third year, and score from the American board of radiology core exam were also tabulated. Pearson correlation coefficients and polynomial regression analysis were used to identify correlations between the total number of interpreted films and clinical, test, and milestone performance.

RESULTS:
Thirty-seven residents interpreted a mean of 12,709 studies (range 8898-19,818; standard deviation [SD] 2351.9) in residency with a mean major discordance rate of 1.1% (range 0.34%- 2.54%; stand dev 0.49%). There was a nonlinear correlation between total number of interpreted films and performance. As the number of interpreted films increased to approximately 16,000, clinical performance (p = 0.004) and test performance (p = 0.01) improved, but volumes over 16,000 correlated with worse performance.

CONCLUSION:
The total number of studies interpreted during radiology training correlates with performance. Residencies should endeavor to find the "sweet spot": the amount of work that maximizes clinical exposure and knowledge without overburdening trainees.

Does Orthopaedic Resident Efficiency Improve with Respect to Decreased Fluoroscopic Times in Tibial Intramedullary Nailing? A Measure of an ACGME Milestone

Bradburn K, Patel JH, Cannada LK. *Current Orthopaedic Practice*. 2019;(2):129. doi:10.1097/ BCO. 0000000000000733.

BACKGROUND:
Intramedullary nailing of tibial fractures is a surgical milestone from the Accreditation Council for Graduate Medical Education (ACGME). Our purpose was to evaluate if fluoroscopic time decreased with increasing resident experience and could be used as a measure of this milestone.

METHODS:
Current Procedural Terminology (CPT) codes were used to identify patients who underwent intramedullary nailing of tibial shaft fractures under the direction of fellowship-trained trauma attending staff. The data collected included patient demographics, fracture classification, fluoroscopic imaging total time, and the post-graduate years (PGY) of orthopaedic residency of the operating resident.

Exclusions of patients included concomitant fluoroscopic procedures, inadequate records, or surgeries involving primary assisting residents with less than PGY-2 experience. We compared overall groups between half years and looked at individual resident years for each of the continuous variables.

RESULTS:
When residents were grouped as senior (PGY-4 and PGY-5) or junior (PGY-2 and PGY-3), seniors used significantly less fluoroscopy than juniors (207.39asec vs. 258.30asec, P=0.018). In the first half of the academic year, PGY-2 residents completed tibial nailing slowest in terms of fluoroscopic usage (P=0.003). PGY-4 residents completed tibial nailing faster in terms of fluoroscopic usage than other years (P=0.031). In the second half of the academic year, PGY-5 residents used significantly less fluoroscopy than PGY-2 residents (P=0.035).

CONCLUSIONS:
As the ACGME currently has no measurement for resident progress and efficiency regarding tibial shaft intramedullary nailing, our data indicate that fluoroscopic measurements may be useful in assessing resident proficiency.

Information within Residency Monthly Evaluation Forms at Two Institutions

ABSTRACT:
Periodic review of resident performance is an important aspect of residency training. Amongst allopathic residency programs, it is expected that the performance of resident physicians which can be grouped based on the ACGME core competencies, be assessed so as to allow for effective feedback and continuous improvement. Review of monthly evaluation forms for residents in the core ACGME programs at Marshall University and the University of Toledo demonstrated a wide spread in the number of Likert questions that faculty were asked to complete. This number ranged from a low of 7 in Surgery to a high of 65 in Psychiatry (both Marshall Programs). Correlation and network analysis were performed on these data. High degrees of correlations were noted between answers to questions (controlled for each resident) on these forms at both institutions. In other words, although evaluation scores varied tremendously amongst the different residents in all the programs studied, scores addressing different competencies tended to be very similar for the same resident, especially in some of the programs which were studied. Network analysis suggested that there were clusters of questions that produced essentially the same answer for a given resident, and these clusters were bigger in some of the different residency program assessment forms. This seemed to be more the rule in the residency programs with large numbers of Likert questions. The authors suggest that reducing the number of monthly questions used to address the core competencies in some programs may be possible without substantial loss of information.

Training, Education for Robotic Performance with Simulation (Terps): A Valuable Tool for Gynecologic Surgeons in Training

Helou C, Seal P, Sanses T, Morozov V, Roque D. Journal of Minimally Invasive Gynecology. 2019;26:S159.

ABSTRACT:
To evaluate the role of robotic simulation in training OBGYN residents by determining an optimal number of exercise repetitions prior to clinical debut; To assess whether clinical exposure accelerates proficiency by correlating laparoscopic/robotic experience with simulator skills acquisition Prospective cohort study Urban academic center with active COEMIG designation 2017-2018 Gynecology residents(PGY1-4) Voluntary participants were instructed to complete 10 repetitions of 5 exercises (pegboard-1, energy dissection-1, energy switching-1, ring&rail-2, tubes) on the dV-Trainer® robotic simulator. After a 4-month hiatus, residents were asked to repeat the protocol. Residents were surveyed regarding prior surgical experience and perceptions regarding simulation utility. 25 of 28(89%) residents participated. Performance was captured using M-scores® (aggregate quality, efficiency, risk, and safety measure). With all exercises, M-scores® increased with repetitions among all levels (mean±SD 58.9±19.1 repetition 1 versus 82.0±13.6 repetition 10, p<0.001); however, after one round, many trainees failed to attain the pre-determined passing score of 80%. Across all participants, mean scores by exercise were 82.5±15.6, 78.0±15.8, 72.6±17.9, 62.7±19.4, 60.1±22.1 (p<0.001). Neither PGY level nor prior surgical experience correlated with higher scores: repetition-1 scores were 61 ±12.8, 54.0±11.2, 59.4 ±19.7, and 59.8±10.6 for PGY-1 through -4 participants, p=0.51; repetition-10 scores were 80±3.9, 82±9.3, 86.5±9.3, and 84.9±9.0, p=0.79, respectively. Self-reported prior surgical experience reflected graduated responsibility: only PGY-4 participants reported console exposure, with most describing 1-5 cases performed. Retention of skills at 4 months negatively correlated with difficulty, suggesting challenging skills require more repetitions to master. Poor compliance hindered data interpretation. The majority of trainees believed simulation is valuable. Robotic simulation may be useful for development/maintenance of robotic skills in Gynecology trainees. M-score® may be insufficiently sensitive; additional metrics should be explored. Robotic simulation is valued by trainees, however, not a milestone established by the ACGME. Protected time with incorporation into curricula would be needed to maximize utility.

Examining the Effects of Narrative Commentary on Evaluators' Summative Assessments of Resident Performance

ABSTRACT:
Anchor-based, end-of-shift ratings are commonly used to conduct performance assessments of resident physicians. These performance evaluations often include narrative assessments, such as solicited or "free-text" commentary. Although narrative commentary can help to create a more detailed and specific assessment of performance, there are limited data describing the effects of narrative commentary on the global assessment process. This single-group, observational study examined the effect of narrative comments on global performance assessments. A subgroup of the clinical competency committee, blinded to resident identity, assigned a single, consensus-based performance score (1-6) to each resident based solely on end-of-shift milestone scores. De-identified narrative comments from end-of-shift evaluations were then included and the process was repeated. We compared milestone-only scores to milestone plus narrative commentary scores using a nonparametric sign test. During the study period, 953 end-of-shift evaluations were submitted on 41 residents. Of these, 535 evaluations included free-text narrative comments. In 17 of the 41 observations, performance scores changed after the addition of narrative comments. In two cases, scores decreased with the addition of free-text commentary. In 15 cases, scores increased. The frequency of net positive change was significant (p = .0023). The addition of narrative commentary to anchor-based ratings significantly influenced the global performance assessment of Emergency Medicine residents by a committee of educators. Descriptive commentary collected at the end of shift may inform more meaningful appraisal of a resident's progress in a milestone-based paradigm. The authors recommend clinical training programs collect unstructured narrative impressions of residents' performance from supervising faculty.

Assessment Methods and Resource Requirements for Milestone Reporting by an Emergency Medicine Clinical Competency Committee

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) introduced milestones for Emergency Medicine (EM) in 2012. Clinical Competency Committees (CCC) are tasked with assessing residents on milestones and reporting them to the ACGME. Appropriate workflows for CCCs are not well defined.

OBJECTIVE:
Our objective was to compare different approaches to milestone assessment by a CCC, quantify resource requirements for each and to identify the most efficient workflow.

DESIGN:
Three distinct processes for rendering milestone assessments were compared: Full milestone assessments (FMA) utilizing all available resident assessment data, Ad-hoc milestone assessments (AMA) created by multiple expert educators using their personal assessment of resident performance, Self-assessments (SMA) completed by residents. FMA were selected as the theoretical gold standard. Intraclass correlation coefficients were used to analyze for agreement between different assessment methods. Kendall's coefficient was used to assess the inter-rater agreement for the AMA.

RESULTS:
All 13 second-year residents and 7 educational faculty of an urban EM Residency Program participated in the study in 2013. Substantial or better agreement between FMA and AMA was seen for 8 of the 23 total subcompetencies (PC4, PC8, PC9, PC11, MK, PROF2, ICS2, SBP2), and for 1 subcompetency (SBP1) between FMA and SMA. Multiple AMA for individual residents demonstrated substantial or better interobserver agreement in 3 subcompetencies (PC1, PC2, and PROF2). FMA took longer to complete compared to AMA (80.9 vs. 5.3 min, p < 0.001).

CONCLUSIONS:
Using AMA to evaluate residents on the milestones takes significantly less time than FMA. However, AMA and SMA agree with FMA on only 8 and 1 subcompetencies, respectively. An estimated 23.5 h of faculty time are required each month to fulfill the requirement for semiannual reporting for a residency with 42 trainees.

Using Operating Room Turnover Time by Anesthesia Trainee Level to Assess Improving Systems-Based Practice Milestones

BACKGROUND:
Operating room (OR) metrics are frequently cited when optimizing cost efficacy and quality of care (Weiss et al, Characteristics of operating room procedures in U.S. hospitals, 2011: Statistical brief #170, 2013; Macario A, Anesthesiology 105:237-240, 2006; Childers et al, JAMA Surg 153:e176233, 2018). Little has been reported to evaluate how anesthesia trainees change anesthesia-related efficiencies in the OR. Statistical correlation may demonstrate awareness and implementation of efficient systems-based practice.

METHODS:
Utilizing computerized OR information systems, specific data regarding anesthesia controlled turnover times were collected (546 data points) over the course of 4 months. The type of surgery performed, patient's American Society of Anesthesiologists (ASA) physical status and OR turnover times were compared for clinical anesthesia (CA) trainee levels CA1, CA2, CA3 and CRNAs. Standard descriptive statistics were computed. Analysis of variance (ANOVA) was performed to compare the average turnover time.

RESULTS:
Average OR turnover time was 31 min ranging from 8 to 60 min. There was a significant difference between the OR turnover time of CA-1 (32 min) compared to CA-3 (29 min) ($p = 0.017$) and CA-1 compared to CRNA (30 min) ($p = 0.016$). OR turnover time was significantly shorter in CA-3 and CRNA. The analysis showed no differences between OR turnover time of ASA categories.

CONCLUSIONS:
These findings posit that trainees improve efficiency over time, but that education may for a time come at the expense of productivity. This trend may demonstrate a more profound understanding and mastery of a learner progressing in the graduate medical education system. This interplay plays a key role in clinical and academic shared success.

Standardized Patients to Assess Resident Interpersonal Communication Skills and Professional Values Milestones

ABSTRACT:
It has been a challenge to assess communication and professional values Milestones in emergency medicine (EM) residents using standardized methods, as mandated by the Accreditation Council for Graduate Medical Education (ACGME). This paper outlines an innovative method of assessing these Milestones using an established instructional method. EM faculty mapped the communication and professional values Milestones to an existing communication and interpersonal skills scale. We identified six communication-focused scenarios: death notification; informed consent; medical non- compliance; medical error; treatment refusal; and advanced directives. In a pilot, 18 EM residents completed these six standardized patient (SP) encounters. Our experience suggests SP encounters can support standardized direct observation of residents' achievement of ACGME Milestones. Further effort can be made to create a tailored, behaviorally-anchored tool that uses the Milestones as the conceptual framework.

Four-Year Analysis of a Novel Milestone-Based Assessment of Faculty by General Surgical Residents

OBJECTIVE:
In response to our faculty's concerns about the quality and reliability of feedback from general surgery residents, we developed a novel faculty assessment tool. This study was designed as an interim analysis of the tool's effectiveness and discriminatory ability.

METHODS:
Our department's educational leadership developed milestones in 7 domains that were scored from 1 to 4, with each level representing an educational approach that ranged from ineffective (1) to ideal (4). Each postgraduate year (PGY) class meets annually to develop a consensus regarding each faculty member's effectiveness in each of the 7 domains: (1) operative supervision, (2) operative teaching, (3) clinic and/or hospital supervision, (4) clinic and/or hospital teaching, (5) conference participation, (6) availability, and (7) overall contribution to the training program. We reviewed the results from the initial 4years of this project. We also analyzed the annual national faculty survey administered by the Accreditation Council for Graduate Medical Education (ACGME) to evaluate faculty satisfaction regarding feedback during the same study period. Data were assessed using the Levene test for homogeneity, analysis of variance, and Wilcoxon-Mann-Whitney tests.

RESULTS:
Forty-two faculty members were annually evaluated by 29 to 32 residents. Each resident PGY class assigned faculty milestone scores that varied across the 7 domains, demonstrating that faculty scores reflected variable opinions about each specific domain, while avoiding labeling an effective faculty member with all high scores and a less effective member with all poor scores. (p< 0.0001). Milestone scores for a given faculty member differed across PGY classes, indicating that junior residents might evaluate a specific faculty member differently than senior residents (p< 0.0001). Eleven faculty members received low scores of 1 or 2 on the overall contribution to training domain and 8/11 (73%) improved to 3 or 4, the following year. Twenty core faculty members were included on the annual ACGME survey. The results from the study period on the ACGME anonymous faculty survey reflected enhanced satisfaction with resident feedback during the study period, improving from 68% to 88% compliance with ACGME standards and our mean program score improved from 4.1 to 4.4 compared to the national mean of 4.3 (p = 0.02).

CONCLUSIONS:
This milestone-based faculty assessment tool improves the quality of the feedback from surgical residents when evaluating faculty. When residents assign a negative statement to describe faculty educational effectiveness in a specific domain, performance improves. A milestone-based faculty assessment strategy should be explored on a national level.

Can Deficiencies in Performance Be Identified Earlier in Surgical Residency? An Initial Report of a Surgical Trainee Assessment of Readiness Exam

OBJECTIVE:
Identifying gaps in medical knowledge, patient management, and procedural competence is difficult early in surgical residency. We designed and implemented an end-of-year examination for our postgraduate year 1 residents, entitled Surgical Trainee Assessment of Readiness (STAR). Our objective in this study was to determine whether STAR scores correlated with other available indicators of resident performance, such as the American Board of Surgery in- training exam (ABSITE) and Milestone scores, and if they provided evidence of additional discriminatory value.

STUDY DESIGN:
Overall and component scores of the STAR exam were compared to the ABSITE and Milestone assessment scores for the 17 categorical residents that took the exam in 2016 and 2017.

SETTING:
Harbor-UCLA Medical Center, a university-affiliated academic medical center.

PARTICIPANTS:
Seventeen categorical general surgery residents.

RESULTS:
The STAR Total Test Score ($\beta = 2.77$, $p = 0.006$) was an independent predictor of the ABSITE taken the same year, and components of the STAR were independent predictors of ABSITE taken the following year. The STAR Total Test Score was lowest in the 3 residents who had at least 1 low Milestone score assessed in the same year; and 2 of these 3 residents had at least 1 low Milestone score assigned the next year after STAR. Lastly, the Patient Care 1 and 2 Milestones assessed in the same year as STAR were uniformly scored as appropriate for level of training, yet the corresponding STAR component for those milestones demonstrated 3 residents as having deficiencies.

CONCLUSIONS:
We have created a multifaceted standardized STAR exam, which correlates with performance on the ABSITE and early milestone scores. It also appears to discriminate resident performance where milestone assessments do not. Further evaluation of the STAR exam with longer term follow-up is needed to confirm these initial findings.

Resident-Specific Morbidity Reduced Following ACS NSQIP Data-Driven Quality Program

BACKGROUND:
The Accreditation Council for Graduate Medical Education Milestone Project for general surgery provided a more robust method for developing and tracking residents' competence. This framework enhanced systematic and progressive development of residents' competencies in surgical quality improvement.

STUDY DESIGN:
A 22-month interactive, educational program based on resident-specific surgical outcomes data culminated in a quality improvement project for postgraduate year 4 surgery residents. Self- assessment, quality knowledge test, and resident-specific American College of Surgeons National Surgical Quality Improvement Program Quality In-Training Initiative morbidity were compared before and after the intervention.

RESULTS:
Quality in-training initiative morbidity decreased from 25% (82/325) to 18% (93/517), p = 0.015 despite residents performing more complex cases. All participants achieved level 4 competency (4/4) within the general surgery milestones improvement of care, practice-based learning and improvement competency. Institutional American College of Surgeons National Surgical Quality Improvement Program general surgery morbidity improved from the ninth to the sixth decile. Quality assessment and improvement self-assessment postintervention scores (M = 23.80, SD = 4.97) were not significantly higher than preintervention scores (M = 19.20, SD = 5.26), p = 0.061. Quality Improvement Knowledge Application Tool postintervention test scores (M = 17.4, SD = 4.88), were not significantly higher than pretest scores (M = 13.2, SD = 1.92), p = 0.12.

CONCLUSION:
Sharing validated resident-specific clinical data with participants was associated with improved surgical outcomes. Participating fourth year surgical residents achieved the highest score, a level 4, in the practice based learning and improvement competency of the improvement of care practice domain and observed significantly reduced surgical morbidity for cases in which they participated.

Predictability of Clinical Knowledge Through Mobile App-based Simulation for the Treatment of Pediatric Septic Arthritis: A Pilot Study

BACKGROUND:
Recently the American Board of Orthopaedic Surgery and the Accreditation Council of Graduate Medical Education have identified the treatment of septic arthritis of the hip in children as a milestone skill for all US orthopaedic residents. The purpose of this study was to test correlation between clinical knowledge and examination score on a mobile app-based training module for the treatment of pediatric septic hip arthritis.

METHODS:
A 4-part simulation model on surgical decision-making associated with the treatment of pediatric septic arthritis was developed through expert consensus. Orthopaedic trainees participating in the "Top Gun" program of the 2015 and 2016 International Pediatric Orthopaedic Symposiums were recruited to participate in this pilot study. Trainees completed a presimulation quiz on their knowledge of diagnosis, arthrocentesis, and surgical irrigation and debridement on a pediatric patient presenting with septic arthritis of the hip. Trainees then completed the 4-part simulation on the mobile app. Pearson correlation analysis was used to assess the relationship between the quiz and the simulation.

RESULTS:
A total of 53 orthopaedic residents and fellows participated in the simulation. Median quiz score was 87 points [interquartile range (IQR), 81 to 94] before the intervention and 100 points (IQR, 94 to 100) postintervention. The median simulation test score was 89 (IQR, 81 to 92) which demonstrated a positive correlation with the postintervention quiz (r=0.44, P<0.001). The preintervention metrics demonstrated a positive correlation with postintervention metrics (r=0.53, P<0.001).

CONCLUSIONS:
This study revealed a statistically significant positive correlation between the mobile app simulation and the clinical knowledge of the participants, as well as the ability to improve knowledge about a procedure during the testing period. These findings support the ability for the mobile app to test clinical knowledge. In the current environment of decreased work hours and patient exposure for orthopaedic trainees, mobile app-based simulation has the potential to safely aid in assessment of orthopaedic residents and fellows.

A Multicenter Collaboration for Simulation-Based Assessment of ACGME Milestones in Emergency Medicine

ABSTRACT:
In 2014, the six allopathic emergency medicine (EM) residency programs in Chicago established an annual, citywide, simulation-based assessment of all postgraduate year 2 EM residents. The cases and corresponding assessment tools were designed by the simulation directors from each of the participating sites. All assessment tools include critical actions that map directly to numerous EM milestones in 11 different subcompetencies. The 2-hour assessments provide opportunities for residents to lead resuscitations of critically ill patients and demonstrate procedural skills, using mannequins and task trainers respectively. More than 80 residents participate annually and their assessment experiences are essentially identical across testing sites. The assessments are completed electronically and comparative performance data are immediately available to program directors.

Radiology Resident Assessment and Feedback Dashboard

ABSTRACT:
Assessment of residents is optimally performed through processes and platforms that provide daily feedback, which can be immediately acted on. Given the documentation required by the Accreditation Council for Graduate Medical Education (ACGME), effective data management, integration, and presentation are crucial to ease the burden of manual documentation and increase the timeliness of actionable information. To this end, the authors modeled the learning activities of residents using the Experience Application Programming Interface (xAPI) framework, which is a standard framework for the learning community. On the basis of the xAPI framework and using open-source software to extend their existing infrastructure, the authors developed a Web-based dashboard that provides residents with a more holistic view of their educational experience. The dashboard was designed around the ACGME radiology milestones and provides real-time feedback to residents using various assessment metrics derived from multiple data sources. The purpose of this article is to describe the dashboard's architecture and components, the design and technical considerations, and the lessons learned in implementing the dashboard.

Individualized Learning Plan (ILP) Is an Effective Tool in Assessing Achievement of Otology- related Subcompetency Milestones

OBJECTIVE:
To investigate the individualized learning plan (ILP) as a tool in assessment of residents' milestone achievements as they pertain to Otology subcompetencies: Chronic Ear Disease, Pediatric Otitis Media, and Hearing Loss.

STUDY DESIGN:
Prospective study.

METHODS:
Twenty otolaryngology residents were instructed to use an ILP and identify six milestones from three otology-related subcompetencies to focus on during the course of a 3-month rotation. They were also asked to plan out specific activities which would help them achieve these milestones, to specify whether or not they successfully achieved them, by what instructional or learning methods and to identify any barriers. The completed ILPs were reviewed by a faculty member.

MAIN OUTCOME MEASURES:
The effectiveness of the ILP was assessed by response compliance rate, corroboration of self- reported milestone achievement with faculty evaluations and the ability to set attainable milestones.

RESULTS:
There was 95% compliance in using an ILP to achieve milestones. Self-reported milestone scores corresponded to the faculty evaluations in a large majority (89.6%) of patients, and tended to be underestimated by the residents. Out of 114 total milestones identified, 44 (38.6%) were not achieved, with particular overestimation in the use of independent study as a learning method.

CONCLUSION:
The ILP is an effective tool in measuring residents' achievement of otology-related milestones, and could possibly be used to supplement or replace faculty assessment. The ILP provides valuable information on barriers to achieving milestones and informs trainees on how to set attainable goals as they pertain to patient care and medical knowledge in otology.

Validity of Simulation-Based Assessment for Accreditation Council for Graduate Medical Education Milestone Achievement

INTRODUCTION:
The Accreditation Council for Graduate Medical Education requires biannual evaluation of anesthesiology residents on 25 subcompetency milestones. Some milestone domains are particularly challenging to repeatedly and reliably observe during clinical care. Simulation-Based Milestones Assessment (SBMA) may help overcome these challenges. However, few studies have examined the external validation of simulation assessment scores (ie, the relationships between simulation-based assessment scores and other standard measures of ability) for milestones. This study analyzed whether SBMA scores (1) discriminate by postgraduate year, (2) improve over time, and (3) correlate with traditional measures of performance.

METHODS:
This is a retrospective analysis of 55 residents' SBMA data from 30 scenarios for two academic years. Each scenario was evaluated for time-in-training discrimination. Scenarios were then analyzed for SBMA scoring trends over time, and SBMA scores were compared with residents' clinical evaluations.

RESULTS:
Twenty-four SBMA scenarios discriminated by postgraduate year. Repeated measure analysis of variance showed statistically significant between-session score improvements ($F (3, 54) = 17.79$, $P < 0.001$). Pearson correlation coefficients demonstrated moderate to strong correlation between SBMA and clinical evaluations: January 2015 $r = 0.67$, $P < 0.01$ (n = 27); May 2015 $r = 0.43$, $P = 0.09$ (n = 17); November 2015 $r = 0.70$, $P < 0.01$ (n = 24); and April 2016 $r = 70$, $P < 0.01$ (n = 27).

CONCLUSIONS:
The associations between SBMA scores and experience level, time-in-training, and clinical performance evaluations provide evidence that SBMA may be used as metrics of residents' Accreditation Council for Graduate Medical Education milestone competencies.

Preliminary Validity Evidence for a Milestones-Based Rating Scale for Chart-Stimulated Recall

BACKGROUND:
Minimally anchored Standard Rating Scales (SRSs), which are widely used in medical education, are hampered by suboptimal interrater reliability. Expert-derived frameworks, such as the Accreditation Council for Graduate Medical Education (ACGME) Milestones, may be helpful in defining level-specific anchors to use on rating scales.

OBJECTIVE:
We examined validity evidence for a Milestones-Based Rating Scale (MBRS) for scoring chart- stimulated recall (CSR).

METHODS:
Two 11-item scoring forms with either an MBRS or SRS were developed. Items and anchors for the MBRS were adapted from the ACGME Internal Medicine Milestones. Six CSR standardized videos were developed. Clinical faculty scored videos using either the MBRS or SRS and following a randomized crossover design. Reliability of the MBRS versus the SRS was compared using intraclass correlation.

RESULTS:
Twenty-two faculty were recruited for instrument testing. Some participants did not complete scoring, leaving a response rate of 15 faculty (7 in the MBRS group and 8 in the SRS group). A total of 529 ratings (number of items × number of scores) using SRSs and 540 using MBRSs were available. Percent agreement was higher for MBRSs for only 2 of 11 items- use of consultants (92 versus 75, P =.019) and unique characteristics of patients (96 versus 79, P =.011)-and the overall score (89 versus 82, P < .001). Interrater agreement was 0.61 for MBRSs and 0.51 for SRSs.

CONCLUSIONS:
Adding milestones to our rating form resulted in significant, but not substantial, improvement in intraclass correlation coefficient. Improvement was inconsistent across items.

Use of a Clinical Pathologic Conference to Demonstrate Residents' ACGME Emergency Medicine Milestones, Aid in Faculty Development, and Increase Academic Output

ABSTRACT:
The Emergency Medicine Milestones Project, developed by the Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Emergency Medicine, includes competence targets for residents to attain and, ultimately, to exceed American Osteopathic Association and ACGME expectations for residents. The authors sought to use the clinical pathologic conference (CPC) format in their institutions' Emergency Medicine Milestones Project to provide measurable residency academic and faculty development outcomes. The CPC is an event in which a resident presents an unknown case to a discussant in advance of a didactic session to demonstrate an organized approach and decision-making rationale to a differential diagnosis. Feedback forms included the assessment of resident discussants from the perspective of level-5 Milestone achievements in particular. Developing an internal CPC competition with a dedicated core faculty coordinator who provides skill development for both resident and faculty presentation has proven successful. Such a competition can document the level-5 achievements for senior residents, be a source of faculty development, and increase peer-reviewed academic output.

Evaluating Surgical Residents' Patient-Centered Communication Skills: Practical Alternatives to the "Apprenticeship Model"

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) requires residency programs to assess communication skills and provide feedback to residents. We aimed to develop a feasible data collection process that generates objective clinical performance information to guide training activities, inform ACGME milestone evaluations, and validate assessment instruments.

DESIGN:
Residents care for patients in the surgical clinic and in the hospital, and participate in a communication curriculum providing practice with standardized patients (SPs). We measured perception of resident communication using the 14-item Communication Assessment Tool (CAT), collecting data from patients at the surgery clinic and surgical wards in the hospital, and from SP encounters during simulated training scenarios. We developed a handout of CAT example behaviors to guide patients completing the communication assessment.

SETTING:
Independent academic medical center.

PARTICIPANTS:
General surgery residents.

RESULTS:
The primary outcome is the percentage of total items patients rated "excellent;" we collected data on 24 of 25 residents. Outpatient evaluations resulted in significantly higher scores (mean 84.5% vs. 68.6%, p < 0.001), and female patients provided nearly statistically significantly higher ratings (mean 85.2% vs. 76.7%, p = 0.084). In multivariate analysis, after controlling for patient gender, visit reason, and race, (1) residents' CAT scores from SPs in simulation were independently associated with communication assessments in their concurrent patient population (p = 0.017), and (2) receiving CAT example instructions was associated with a lower percentage of excellent ratings by 9.3% (p = 0.047).

CONCLUSIONS:
Our data collection process provides a model for obtaining meaningful information about resident communication proficiency. CAT evaluations of surgical residents by the inpatient population had not previously been described in the literature; our results provide important insight into relationships between the evaluations provided by inpatients, clinic patients, and SPs in simulation. Our example behaviors guide shows promise for addressing a common concern, minimizing ceiling effects when measuring physician-patient communication.

Teaching and Assessing Professionalism in Radiology: Resources and Scholarly Opportunities to Contribute to Required Expectations

ABSTRACT:
Teaching and assessing trainees' professionalism now represents an explicit expectation for Accreditation Council Graduate Medical Education-accredited radiology programs. Challenges to meeting this expectation include variability in defining the construct of professionalism; limits of traditional teaching and assessment methods, used for competencies historically more prominent in medical education, for professionalism; and emerging expectations for credible and feasible professionalism teaching and assessment practices in the current context of health care training and practice. This article identifies promising teaching resources and methods that can be used strategically to augment traditional teaching of the cognitive basis for professionalism, including role modeling, case- based scenarios, debriefing, simulations, narrative medicine (storytelling), guided discussions, peer- assisted learning, and reflective practice. This article also summarizes assessment practices intended to promote learning, as well as to inform how and when to assess trainees as their professional identities develop over time, settings, and autonomous practice, particularly in terms of measurable behaviors. This includes assessment tools (including mini observations, critical incident reports, and appreciative inquiry) for authentic assessment in the workplace; engaging multiple sources (self-peer, other health professionals, and patients) in assessment; and intentional practices for trainees to take responsibility for seeking our actionable feedback and reflection. This article examines the emerging evidence of the feasibility and value added of assessment of medical competency milestones, including professionalism, coordinated by the Accreditation Council Graduate Medical Education in radiology and other medical specialties. Radiology has a strategic opportunity to contribute to scholarship and inform policies in professionalism teaching and assessment practices.

Resident and Program Director's Perceptions of Milestone-Based Feedback in Obstetrics and Gynecology

INTRODUCTION:
In July 2014, US residency programs fully implemented the Next Accreditation System including the use of milestone evaluation and reporting. Currently, there has been little investigation into the result of implementation of this new system. Therefore, this study sought to evaluate perceptions of Obstetrics and Gynecology residents and program directors regarding the use of milestone-based feedback and identify areas of deficiency.

METHODS:
A Web-based survey was sent to US-based Obstetrics and Gynecology residents and program directors regarding milestone-based assessment implementation.

RESULTS:
Out of 245 program directors, 84 responded to our survey (34.3% response rate). Of responding program directors, most reported that milestone-based feedback was useful (74.7%), fair (83.0%), and accurate (76.5%); however, they found it administratively burdensome (78.1%). Residents felt that milestone-based feedback was useful (62.7%) and fair (70.0%). About 64.3% of residents and 74.7% of program directors stated that milestone-based feedback is an effective tool to track resident progression; however, a sizable minority of both groups believe that it does not capture surgical aptitude. Qualitative analysis of free response comments was largely negative and highlighted the administrative burden and lack of accuracy of milestone- based feedback.

CONCLUTION:
Overall, both Obstetrics and Gynecology program directors and residents report that milestone- based feedback is useful and fair. Issues of administrative burden, timeliness, evaluation of surgical aptitude, and ability to act on assigned milestone levels were identified. Although this study is limited to one specialty, such issues are likely important to all residents, faculty, and program directors who have implemented the Next Accreditation System requirements.

Shadowing Emergency Medicine Residents by Medical Education Specialists to Provide Feedback on Non-Medical Knowledge-Based ACGME Sub-Competencies

OBJECTIVE:
Non-medical knowledge-based sub-competencies (multitasking, professionalism, accountability, patient-centered communication, and team management) are challenging for a supervising emergency medicine (EM) physician to evaluate in real-time on shift while also managing a busy emergency department (ED). This study examines residents' perceptions of having a medical education specialist shadow and evaluate their nonmedical knowledge skills.

METHODS:
Medical education specialists shadowed postgraduate year 1 and postgraduate year 2 EM residents during an ED shift once per academic year. In an attempt to increase meaningful feedback to the residents, these specialists evaluated resident performance in selected non- medical knowledge-based Accreditation Council of Graduate Medical Education (ACGME) sub- competencies and provided residents with direct, real-time feedback, followed by a written evaluation sent via email. Evaluations provided specific references to examples of behaviors observed during the shift and connected these back to ACGME competencies and milestones.

RESULTS:
Twelve residents participated in this shadow experience (six post graduate year 1 and six postgraduate year 2). Two residents emailed the medical education specialists ahead of the scheduled shadow shift requesting specific feedback. When queried, five residents voluntarily requested their feedback to be included in their formal biannual review. Residents received milestone scores and narrative feedback on the non-medical knowledge-based ACGME sub-competencies and indicated the shadow experience and subsequent feedback were valuable.

CONCLUSION:
Medical education specialists who observe residents over the course of an entire shift and evaluate non-medical knowledge-based skills are perceived by EM residents to provide meaningful feedback and add valuable information for the biannual review process.

Patient-Centered Medical Home Status and Preparedness to Assess Resident Milestones: A CERA Study

PURPOSE:
The patient-centered medical home (PCMH) model has been proposed as the ideal model for delivering primary care and is focused on improving patient safety and quality, reducing costs, and enhancing patient satisfaction. The mandated Accreditation Council for Graduate Medical Education educational milestones for evaluation of resident competency represent the skills graduates will utilize after graduation. Many of these skills are reflected in the PCMH model. We sought to determine if residency programs whose main family medicine (FM) practice sites have achieved PCMH recognition are therefore more prepared to evaluate milestones.

METHOD:
A national Council of Academic Family Medicine Educational Research Alliance (CERA) survey of family medicine program directors (PDs) was conducted during June and July 2015 to determine if PCMH recognition influences PDs' ability to evaluate training methods and their level of preparedness to evaluate milestones.

RESULTS:
The response rate for the survey was 53.3% (252/473). Nearly two-thirds of the PDs (62.7%) reported that their main FM practice site had earned PCMH recognition. There was no statistical difference between non-PCMH-recognized vs PCMH-recognized programs in how PDs perceived that their program was prepared to assess residents' milestone levels overall (P=0.414). Residents of PCMH- recognized programs were more likely to receive training for team-based care (P=0.009), system improvement plans (P<0.001), root-cause analysis (P=0.002), and health behavior change (P=0.003).

CONCLUSIONS:
PCMH recognition itself did not improve preparedness of FM residency programs to assess milestones. Residents from programs whose main FM practice site is PCMH-recognized are more likely to be trained in the key concepts and tasks associated with the PCMH model, tools that they are expected to utilize extensively after graduation.

Comparing Real-Time versus Delayed Video Assessments for Evaluating ACGME Sub- Competency Milestones in Simulated Patient Care Environments

BACKGROUND:
Background Simulation is an effective method for creating objective summative assessments of resident trainees. Real-time assessment (RTA) in simulated patient care environments is logistically challenging, especially when evaluating a large group of residents in multiple simulation scenarios. To date, there is very little data comparing RTA with delayed (hours, days, or weeks later) video-based assessment (DA) for simulation-based assessments of Accreditation Council for Graduate Medical Education (ACGME) sub-competency milestones. We hypothesized that sub-competency milestone evaluation scores obtained from DA, via audio-video recordings, are equivalent to the scores obtained from RTA.

METHODS:
Forty-one anesthesiology residents were evaluated in three separate simulated scenarios, representing different ACGME sub-competency milestones. All scenarios had one faculty member perform RTA and two additional faculty members perform DA. Subsequently, the scores generated by RTA were compared with the average scores generated by DA. Variance component analysis was conducted to assess the amount of variation in scores attributable to residents and raters.

RESULTS:
Paired t-tests showed no significant difference in scores between RTA and averaged DA for all cases. Cases 1, 2, and 3 showed an intraclass correlation coefficient (ICC) of 0.67, 0.85, and 0.50 for agreement between RTA scores and averaged DA scores, respectively. Analysis of variance of the scores assigned by the three raters showed a small proportion of variance attributable to raters (4% to 15%).

CONCLUSIONS:
The results demonstrate that video-based delayed assessment is as reliable as real-time assessment, as both assessment methods yielded comparable scores. Based on a department's needs or logistical constraints, our findings support the use of either real-time or delayed video evaluation for assessing milestones in a simulated patient care environment.

Simulation for Assessment of Milestones in Emergency Medicine Residents

OBJECTIVES:
All residency programs in the United States are required to report their residents' progress on the milestones to the Accreditation Council for Graduate Medical Education (ACGME) biannually. Since the development and institution of this competency-based assessment framework, residency programs have been attempting to ascertain the best ways to assess resident performance on these metrics. Simulation was recommended by the ACGME as one method of assessment for many of the milestone subcompetencies. We developed three simulation scenarios with scenario-specific milestone-based assessment tools. We aimed to gather validity evidence for this tool.

METHODS:
We conducted a prospective observational study to investigate the validity evidence for three mannequin-based simulation scenarios for assessing individual residents on emergency medicine (EM) milestones. The subcompetencies (i.e., patient care [PC]1, PC2, PC3) included were identified via a modified Delphi technique using a group of experienced EM simulationists. The scenario-specific checklist (CL) items were designed based on the individual milestone items within each EM subcompetency chosen for assessment and reviewed by experienced EM simulationists. Two independent live raters who were EM faculty at the respective study sites scored each scenario following brief rater training. The inter-rater reliability (IRR) of the assessment tool was determined by measuring intraclass correlation coefficient (ICC) for the sum of the CL items as well as the global rating scales (GRSs) for each scenario. Comparing GRS and CL scores between various postgraduate year (PGY) levels was performed with analysis of variance.

RESULTS:
Eight subcompetencies were chosen to assess with three simulation cases, using 118 subjects. Evidence of test content, internal structure, response process, and relations with other variables were found. The ICCs for the sum of the CL items and the GRSs were >0.8 for all cases, with one exception (clinical management GRS = 0.74 in sepsis case). The sum of CL items and GRSs ($p < 0.05$) discriminated between PGY levels on all cases. However, when the specific CL items were mapped back to milestones in various proficiency levels, the milestones in the higher proficiency levels (level 3 [L3] and 4 [L4]) did not often discriminate between various PGY levels. L3 milestone items discriminated between PGY levels on five of 12 occasions they were assessed, and L4 items discriminated only two of 12 times they were assessed.

CONCLUSION:
Three simulation cases with scenario-specific assessment tools allowed evaluation of EM residents on proficiency L1 to L4 within eight of the EM milestone subcompetencies. Evidence of test content, internal structure, response process, and relations with other variables were found. Good to excellent IRR and the ability to discriminate between various PGY levels was found for both the sum of CL items and the GRSs. However, there was a lack of a positive relationship between advancing PGY level and the completion of higher-level milestone items (L3 and L4).

myTIPreport and Training for Independent Practice: A Tool for Real-Time Workplace Feedback for Milestones and Procedural Skills

BACKGROUND:
Few tools currently exist for effective, accessible delivery of real-time, workplace feedback in the clinical setting.

OBJECTIVE:
We developed and implemented a real-time, web-based tool for performance-based feedback in the clinical environment.

METHODS:
The tool (myTIPreport) was designed for performance-based feedback to learners on the Accreditation Council for Graduate Medical Education (ACGME) Milestones and procedural skills. "TIP" stands for "Training for Independent Practice." We implemented myTIPreport in obstetrics and gynecology (Ob-Gyn) and female pelvic medicine and reconstructive surgery (FPMRS) programs between November 2014 and May 2015. Residents, fellows, teachers, and program directors completed preimplementation and postimplementation surveys on their perceptions of feedback.

RESULTS:
Preimplementation surveys were completed by 656 participants of a total of 980 learners and teachers in 19 programs (12 Ob-Gyn and 7 FPMRS). This represented 72% (273 of 378) of learners and 64% (383 of 602) of teachers. Seventy percent of participants (381 of 546) reported having their own *individual processes* for real-time feedback; the majority (79%, 340 of 430) described these processes as *informal discussions*. Over 6 months, one-third of teachers and two-thirds of learners used the myTIPreport tool a total of 4311 times. Milestone feedback was recorded 944 times, and procedural feedback was recorded 3367 times. Feedback addressed all ACGME Milestones and procedures programmed into myTIPreport. Most program directors reported that tool implementation was successful.

CONCLUSIONS:
The majority of learners successfully received workplace feedback using myTIPreport. This web-based tool, incorporating procedures and ACGME Milestones, may be an important transition from other feedback formats.

Programmatic Assessment in Emergency Medicine: Implementation of Best Practices

BACKGROUND:
Programmatic assessment is the intentional collection of key data from multiple sources for both *assessment of learning* and *assessment for learning*.

OBJECTIVE:
We developed a system of programmatic assessment (PA) to identify competency progression (summative) and assessment for learning to assist residents in their formative development.

METHODS:
The programmatic assessment was designed iteratively from 2014 through 2016. All assessments were first categorized by competency domain and source of assessment. The number of assessment modalities for each competency domain was collected. These multisource assessments were then mapped by program leadership to the milestones to develop a master PA blueprint. A resident learning management system provided the platform for aggregating formative and summative data, allowing residents and faculty ongoing access to guide learning and assessment. A key component of programmatic assessment was to support resident integration of assessment information through feedback by faculty after shifts and during monthly formal assessments, semiannual resident reviews, and summative judgments by the Clinical Competency Committee.

RESULTS:
Through the PA, the 6 competency domains are assessed through multiple modalities: patient care (22 different assessments), professionalism (18), systems-based practice (17), interprofessional and communication skills (16), medical knowledge (11), and practice-based learning and improvement (6). Each assessment provides feedback to the resident in various formats. Our programmatic assessment has been utilized for more than 2 years with iterative improvements.

CONCLUSIONS:
The implementation of programmatic assessment allowed our program to organize diverse, multisourced feedback to drive both formative and summative assessments.

Do End-of-Rotation and End-of-Shift Assessments Inform Clinical Competency Committees' (CCC) Decisions?

INTRODUCTION:
Clinical Competency Committees (CCC) require reliable, objective data to inform decisions regarding assignment of milestone proficiency levels, which must be reported to the Accreditation Council for Graduate Medical Education. After the development of two new assessment methods, the end-of-shift (EOS) assessment and the end-of-rotation (EOR) assessment, we sought to evaluate their performance. We report data on the concordance between these assessments, as well as how each informs the final proficiency level determined in biannual CCC meetings. We hypothesized that there would be a high concordance level between the two assessment methods, including concordance of both the EOS and EOR with the final proficiency level designation by the CCC.

METHODS:
The residency program is an urban academic four-year emergency medicine residency with 48 residents. After their shifts in the emergency department (ED), residents handed out EOS assessment forms asking about individual milestones from 15 subcompetencies to supervising physicians, as well as triggered electronic EOR-doctor (EORd) assessments to supervising doctors and EOR-nurse (EORn) to nurses they had worked with after each two-week ED block. EORd assessments contained the full proficiency level scale from 16 subcompetencies, while EORn assessments contained four subcompetencies. Data reports were generated after each six-month assessment period and data was aggregated. We calculated Spearman's rank order correlations for correlations between assessment types and between assessments and final CCC proficiency levels.

RESULTS:
Over 24 months, 5,234 assessments were completed. The strongest correlations with CCC proficiency levels were the EORd for the immediate six-month assessment period prior (r_s 0.71- 0.84), and the CCC proficiency levels from the previous six-months (r_s 0.83-0.92). EOS assessments had weaker correlations (r_s 0.49 to 0.62), as did EORn (r_s 0.4 to 0.73).

CONCLUSION:
End-of-rotation assessments completed by supervising doctors are most highly correlated with final CCC proficiency level designations, while end-of-shift assessments and end-of-rotation assessments by nurses did not correlate strongly with final CCC proficiency levels, both with overestimation of levels noted. Every level of proficiency the CCC assigned appears to be highly correlated with the designated level in the immediate six-month period, perhaps implying CCC members are biased by previous level assignments.

Straight Line Scoring by Clinical Competency Committees Using Emergency Medicine Milestones

Beeson MS, Hamstra SJ, Barton MA, Yamazaki K, Counselman FL, Shayne PH, Holmboe ES, Muelleman RL, Reisdorff EJ. J Grad Med Educ. 2017 Dec;9(6):716-720. doi: 10.4300/JGME-D- 17- 00304.1.

BACKGROUND:
In 2013, milestone ratings became a reporting requirement for emergency medicine (EM) residency programs. Programs rate each resident in the fall and spring on 23 milestone subcompetencies.

OBJECTIVE:
This study examined the incidence of straight line scoring (SLS) for EM Milestone ratings, defined as a resident being assessed the same score across the milestone subcompetencies.

METHODS:
This descriptive analysis measured the frequencies of SLS for all Accreditation Council for Graduate Medical Education (ACGME)-accredited EM programs during the 2015-2016 academic year. Outcomes were the frequency of SLS in the fall and spring milestone assessments, changes in the number of SLS reports, and reporting trends. Chi-square analysis compared nominal variables.

RESULTS:
There were 6257 residents in the fall and 6588 in the spring. Milestone scores were reported for 6173 EM residents in the fall (99% of 6257) and spring (94% of 6588). In the fall, 93% (5753 residents) did not receive SLS ratings and 420 (7%) did, with no significant difference compared with the spring (5776 [94%] versus 397 [6%]). Subgroup analysis showed higher SLS results for residents' first ratings (183 of 2136 versus 237 of 4220, $P < .0001$) and for their final ratings (200 of 2019 versus 197 of 4354, $P < .0001$). Twenty percent of programs submitted 10% or more SLS ratings, and a small percentage submitted more than 50% of ratings as SLS.

CONCLUSIONS:
Most programs did not submit SLS ratings. Because of the statistical improbability of SLS, any SLS ratings reduce the validity assertions of the milestone assessments.

Comprehensive Assessment of Struggling Learners Referred to a Graduate Medical Education Remediation Program

BACKGROUND:
Implementation of the Next Accreditation System has provided a standardized framework for identifying learners not meeting milestones, but there is as yet no corresponding framework for remediation.

OBJECTIVE:
We developed a comprehensive assessment process that allows correct diagnosis of a struggling learner's deficit(s) to promote successful remediation.

METHODS:
At the University of Pennsylvania, resident learners within the Department of Medicine who are not meeting milestones are referred to the Early Intervention Remediation Committee (EIRC). The EIRC, composed of 14 faculty members with expertise in remediation, uses a standardized process to assess learners' deficits. These faculty members categorize primary deficits as follows: medical knowledge, clinical reasoning, organization and efficiency, professionalism, and communication skills. The standardized process of assessment includes an analysis of the learner's file, direct communication with evaluators, an interview focused on learner perception of the problem, screening for underlying medical or psychosocial issues, and a review of systems for deficits in the 6 core competencies.
Participants were surveyed after participating in this process.

RESULTS:
Over a 2-year period, the EIRC assessed and developed remediation plans for 4% of learners (14 of a total 342). Following remediation and reassessment, the identified problems were satisfactorily resolved in all cases with no disciplinary action. While the process was time intensive, an average of 45 hours per learner, the majority of faculty and residents rated it as positive and beneficial.

CONCLUSIONS:
This structured assessment process identifies targeted areas for remediation and adds to the tools available to Clinical Competency Committees.

Competency Assessment in Family Medicine Residency: Observations, Knowledge-Based Examinations, and Advancement

BACKGROUND:
The Family Medicine (FM) Milestones are competency-based assessments of residents in key dimensions relevant to practice in the specialty. Residency programs use the milestones in semiannual reviews of resident performance from the time of entry into the program to graduation.

OBJECTIVE:
Using a national sample, we investigated the relationship of FM competency-based assessments to resident progress and the complementarity of milestones with knowledge-based assessments in FM residencies.

METHODS:
We used midyear and end-of-year milestone ratings for all FM residents in Accreditation Council for Graduate Medical Education-accredited programs during academic years 2014-2015 and 2015-2016. The milestones contain 22 items across 6 competencies. We created a summative index across the milestones. The American Board of Family Medicine database provided resident demographics and in- training examination (ITE) scores. We linked information to the milestone data.

RESULTS:
The sample encompassed 6630 FM residents. The summative milestone index increased, on average, for each cohort (postgraduate year 1 [PGY-1] to PGY-2 and PGY-2 to PGY-3) at each assessment. The correlation between the milestone index that excluded the medical knowledge milestone and ITE scores was $r = .195$ ($P < .001$) for PGY-1 to PGY-2 cohort and $r = .254$ ($P < .001$) for PGY-2 to PGY-3 cohort. For both cohorts, ITE scores and composite milestone assessments were higher for residents who advanced than for those who did not.

CONCLUSIONS:
Competency-based assessment using the milestones for FM residents seems to be a viable multidimensional tool to assess the successful progression of residents.

Two-Year Experience Implementing a Curriculum to Improve Residents' Patient-Centered Communication Skills

OBJECTIVES:
Surgery milestones from The Accreditation Council for Graduate Medical Education have encouraged a focus on training and assessment of residents' nontechnical skills, including communication. We describe our 2-year experience implementing a simulation-based curriculum, results of annual communication performance assessments, and resident evaluations.

DESIGN:
Eight quarterly modules were conducted on various communication topics. Former patient volunteers served as simulation participants (SP) who completed annual assessments using the Communication Assessment Tool (CAT). During these 2 modules, communication skills were assessed in the following standardized scenarios: (1) delivering bad news to a caregiver of a patient with postoperative intracerebral hemorrhage and (2) primary care gallstone referral with contraindications for cholecystectomy. SP-CAT ratings were evaluated for correlations by individual and associations with trainee and SP characteristics. Surgical patient experience surveys are evaluated during the curriculum.

SETTING:
Independent academic medical center surgical simulation center.

PARTICIPANTS:
Twenty-five surgery residents per year in 2015 to 2017.

RESULTS:
Residents have practiced skills in a variety of scenarios including bad news delivery, medical error disclosure, empathic communication, and end-of-life conversations. Residents report positive learning experiences from the curriculum (90% graded all modules A/A+). Confidence ratings rose following each module ($p < 0.001$) and in the second year ($p < 0.001$). Annual assessments yielded insights into skills level, and relationships to resident confidence levels and traits. Communication scores were not associated with resident gender or postgraduate year. Over the course of the curriculum implementation, surgical patients have reported that doctors provided explanations with improved clarity ($p = 0.042$).

CONCLUSIONS:
The simulation-based SP-CAT has shown initial evidence of usability, content validity, relationships to observed communication behaviors and residents' skills confidence. Evaluations of different scenarios may not be correlated for individuals over time. The communication curriculum paralleled improvements in patient experience concerning surgeons' clear explanations. An ongoing surgery resident communication curriculum has numerous educational, assessment, and institutional benefits.

Impact of an Event Reporting System on Resident Complication Reporting in Plastic Surgery Training: Addressing an ACGME and Plastic Surgery Milestone Project Core Competency

BACKGROUND:
The Accreditation Council for Graduate Medical Education and Plastic Surgery Milestone Project has identified practice-based learning and improvement, which involves systematically analyzing current practices and implementing changes, as a core competency in residency education. In surgical care, complication reporting is an essential component of practice-based learning and improvement as complications are analyzed in morbidity and mortality conference for quality improvement. Unfortunately, current methods for capturing a comprehensive profile of complications may significantly underestimate the true occurrence of complications. Therefore, the objectives of this study are to evaluate an intervention for complication reporting and compare this to current practice, in a plastic surgery training program.

METHODS:
This is a preintervention and postintervention study evaluating resident reporting of complications on a plastic surgery service. The intervention was an online event reporting system developed by department leadership and patient safety experts. The cohorts consisted of all patients undergoing surgery during two separate 3-month blocks bridged by an implementation period. A trained reviewer recorded complications, and this served as the reference standard. Fisher's exact test was used for binary comparisons.

RESULTS:
There were 32 complications detected in 219 patients from June to August of 2015 and 35 complications in 202 patients from October to December of 2015. The proportion of complications reported in the preintervention group was nine of 32 (28.1 percent). After the intervention, this significantly increased to 32 of 35 (91.4 percent) (p < 0.001).

CONCLUSION:
An intervention using an event reporting system, supported by departmental leadership, led to significant improvements in complication reporting by plastic surgery residents.

Implementation of a Service-Specific Template Integrating Objective Structured Clinical Examination and Accreditation Council for Graduate Medical Education Milestones: One Institution's Experience

ABSTRACT:
The Accreditation Council for Graduate Medical Education (ACGME) orthopaedic milestones require detailed, frequent resident evaluations. This institution desired a cost-effective objective structured clinical examination (OSCE) to facilitate these evaluations. Data were collected as a prospective, uncontrolled observational study. The OSCE was completed by residents entering and exiting the foot and ankle rotation during postgraduate years 2 and 4. Physician assistants functioned as standardized patients. Statistical analyses were performed using paired and independent t tests. The OSCE was implemented using reliable, low-cost modalities and has facilitated milestones evaluations. Preliminary data show 4th-year residents performed higher in prerotation global assessment with a standardized patient and written exam (p < .03). Second- year residents showed improvement in the written exam on rotation completion (p = .03). Using this methodology, institutions may establish similar cost-effective OSCEs as feasible evaluative solutions to satisfy milestone requirements. The authors believe this tool may be modified for any specialty. (Journal of Surgical Orthopaedic Advances 26(4):257-261, 2017).

A Checklist to Help Faculty Assess ACGME Milestones in a Video-Recorded OSCE

BACKGROUND:
Faculty members need to assess resident performance using the Accreditation Council for Graduate Medical Education Milestones.

OBJECTIVE:
In this randomized study we used an objective structured clinical examination (OSCE) around the disclosure of an adverse event to determine whether use of a checklist improved the quality of milestone assessments by faculty.

METHODS:
In 2013, a total of 20 anesthesiology faculty members from 3 institutions were randomized to 2 groups to assess 5 videos of trainees demonstrating advancing levels of competency on the OSCE. One group used milestones alone, and the other used milestones plus a 13-item checklist with behavioral anchors based on ideal performance. We classified faculty ratings as either correct or incorrect with regard to the competency level demonstrated in each video, and then used logistic regression analysis to assess the effect of checklist use on the odds of correct classification.

RESULTS:
Thirteen of 20 faculty members rated assessing performance using milestones alone as *difficult* or *very difficult*. Checklist use was associated with significantly greater odds of correct classification at entry level (odds ratio [OR] = 9.2, 95% confidence interval [CI] 4.0-21.2) and at junior level (OR = 2.7, 95% CI 1.3-5.7) performance. For performance at other competency levels checklist use did not affect the odds of correct classification.

CONCLUSIONS:
A majority of anesthesiology faculty members reported difficulty with assessing a videotaped OSCE of error disclosure using milestones as primary assessment tools. Use of the checklist assisted in correct assessments at the entry and junior levels.

Development of a Global Health Milestones Tool for Learners in Emergency Medicine: A Pilot Project

OBJECTIVES:
In medical education and training, increasing numbers of institutions and learners are participating in global health experiences. Within the context of competency-based education and assessment methodologies, a standardized assessment tool may prove valuable to all of the aforementioned stakeholders. Milestones are now used as the standard for trainee assessment in graduate medical education. Thus, the development of a similar, milestone- based tool was undertaken, with learners in emergency medicine (EM) and global health in mind.

METHODS:
The Global Emergency Medicine Think Tank Education Working Group convened at the 2016 Society for Academic Medicine Annual Meeting in New Orleans, Louisiana. Using the Interprofessional Global Health Competencies published by the Consortium of Universities for Global Health's Education Committee as a foundation, the working group developed individual milestones based on the 11 stated domains. An iterative review process was implemented by teams focused on each domain to develop a final product.

RESULTS:
Milestones were developed in each of the 11 domains, with five competency levels for each domain. Specific learning resources were identified for each competency level and assessment methodologies were aligned with the milestones framework. The Global Health Milestones Tool for learners in EM is designed for continuous usage by learners and mentors across a career.

CONCLUSIONS:
This Global Health Milestones Tool for learners in EM may prove valuable to numerous stakeholders. The next steps include a formalized pilot program for testing the tool's validity and usability across training programs, as well as an assessment of perceived utility and applicability by collaborating colleagues working in training sites abroad.

Mapping Residency Global Health Experiences to the ACGME Family Medicine Milestones

Grissom MO, Iroku-Malize T, Peila R, Perez M, Philippe N. Fam Med. 2017 Jul;49(7):553-557.

BACKGROUND AND OBJECTIVES:
Global health (GH) experiences are a unique part of family medicine (FM) training that offer an opportunity for residents to demonstrate development across a multitude of the milestones recently implemented by the Accreditation Council for Graduate Medical Education (ACGME). The GH experience presents an opportunity for resident development, and including a component of written reflection can provide tangible evidence of development in areas that can be difficult to assess.

METHODS:
A mixed methods approach was used to integrate quantitative (frequency) data with qualitative content from the written reflections of 12 of our FM residents who participated in GH experiences.

RESULTS:
Written reflections touched on each of the 22 milestones, although some milestones were noted more frequently than others. The most commonly identified milestones fell within the competency areas of systems-based practice, professionalism, and practice-based learning and improvement. Our qualitative approach allowed us to gain an appreciation of the unique experiences that demonstrated growth across the various milestones.

CONCLUSIONS:
We conclude that any program that offers GH experiences should incorporate some form of written reflection to maximize resident growth and offer evaluative faculty a window into that development.

The Impact of Using Mean versus Mode When Assessing Resident Competency

BACKGROUND:
The Accreditation Council for Graduate Medical Education Milestone Project was implemented in 2014 to standardize assessments and progression of residents. While it is recommended that milestones not be used as tools for direct assessments of resident competency, many programs have used or adapted milestone tools for this purpose.

OBJECTIVE:
We sought to explore use of the most frequent milestone level at which a resident was evaluated (ie, the mode), and compared this to the standard practice of using the arithmetic mean for summarizing performance.

METHODS:
We reviewed all Family Medicine Milestone evaluations from 1 program for the first 2 academic years of milestone implementation. Mean and mode scores were calculated across 24 unique residents, 841 evaluation forms, and 5897 measurements. The proportion of overestimation errors (where the mean is at least 0.5 larger than the mode) and underestimation errors (where the mean is at least 0.5 less than the mode) were then compared across resident training year and subcompetency.

RESULTS:
For the 24 residents, an estimation error occurred in 175 of 792 of the comparisons (22%). Of these errors, 118 (67%) were overestimation errors. First-year residents accounted for 55% (96 of 175) of all estimation errors. All subcompetencies had some estimation errors, with 6 having greater than 5%.

CONCLUSIONS:
If the trend for using the milestones as stand-alone assessment tools is to continue, aggregating data by using frequency distributions and mode would be a more stable and appropriate approach given their nominal or, at best, ordinal nature.

Rapid Web-Based Platform for Assessment of Orthopedic Surgery Patient Care Milestones: A 2- Year Validation

Gundle KR, Mickelson DT, Cherones A, Black J, Hanel DP. J Surg Educ. 2017 May 18. pii: S1931- 7204(16) 30242-2. doi: 10.1016/j.jsurg.2017.05.001.

OBJECTIVE:
To determine the validity, feasibility, and responsiveness of a new web-based platform for rapid milestone-based evaluations of orthopedic surgery residents.

SETTING:
Single academic medical center, including a trauma center and pediatrics tertiary hospital.

PARTICIPANTS:
Forty residents (PG1-5) in an orthopedic residency program and their faculty evaluators.

METHODS:
Residents and faculty were trained and supported in the use of a novel trainee-initiated web- based evaluation system. Residents were encouraged to use the system to track progress on patient care subcompetencies. Two years of prospectively collected data were reviewed from residents at an academic program. The primary outcome was Spearman's rank correlation between postgraduate year (PGY) and competency level achieved as a measure of validity. Secondary outcomes assessed feasibility, resident self-evaluation versus faculty evaluation, the distributions among subcompetencies, and responsiveness over time.

RESULTS:
Between February 2014 and February 2016, 856 orthopedic surgery patient care subcompetency evaluations were completed (1.2 evaluations per day). Residents promptly requested feedback after a procedure (median = 0 days, interquartile range: 0-2), and faculty responded within 2 days in 51% (median = 2 days, interquartile range: 0-13). Primary outcome showed a correlation between PGY and competency level ($r = 0.78$, $p < 0.001$), with significant differences in competency among PGYs ($p < 0.001$ by Kruskal-Wallis rank sum test). Self- evaluations by residents substantially agreed with faculty- assigned competency level (weighted Cohen's K = 0.72, $p < 0.001$). Resident classes beginning the study as PGY1, 2, and 3 separately demonstrated gains in competency over time (Spearman's rank correlation 0.39, 0.60, 0.59, respectively, each $p < 0.001$). There was significant variance in the number of evaluations submitted per subcompetency (median = 43, range: 6-113) and competency level assigned ($p < 0.01$).

CONCLUSIONS:
Rapid tracking of trainee competency with milestone-based evaluations in a learner-centered mobile platform demonstrated validity, feasibility, and responsiveness. Next Accreditation System-mandated data may be efficiently collected and used for trainee and program self-study.

Comparison of Male vs Female Resident Milestone Evaluations by Faculty during Emergency Medicine Residency Training

Dayal A, O'Connor DM, Qadri U, Arora VM. JAMA Intern Med. 2017 May 1;177(5):651-657.

IMPORTANCE:
Although implicit bias in medical training has long been suspected, it has been difficult to study using objective measures, and the influence of sex and gender in the evaluation of medical trainees is unknown. The emergency medicine (EM) milestones provide a standardized framework for longitudinal resident assessment, allowing for analysis of resident performance across all years and programs at a scope and level of detail never previously possible.

OBJECTIVE:
To compare faculty-observed training milestone attainment of male vs female residency training.

DESIGN, SETTING, AND PARTICPANTS:
This multicenter, longitudinal, retrospective cohort study took place at 8 community and academic EM training programs across the United States from July 1, 2013, to July 1, 2015, using a real-time, mobile- based, direct-observation evaluation tool. The study examined 33 456 direct-observation subcompetency evaluations of 359 EM residents by 285 faculty members.

MAIN OUTCOMES AND MEASURES:
Milestone attainment for male and female EM residents as observed by male and female faculty throughout residency and analyzed using multilevel mixed-effects linear regression modeling.

RESULTS:
A total of 33 456 direct-observation evaluations were collected from 359 EM residents (237 men [66.0%] and 122 women [34.0%]) by 285 faculty members (194 men [68.1%] and 91 women [31.9%]) during the study period. Female and male residents achieved similar milestone levels during the first year of residency. However, the rate of milestone attainment was 12.7% (0.07 levels per year) higher for male residents through all of residency (95% CI, 0.04-0.09). By graduation, men scored approximately 0.15 milestone levels higher than women, which is equivalent to 3 to 4 months of additional training, given that the average resident gains approximately 0.52 levels per year using our model (95% CI, 0.49-0.54). No statistically significant differences in scores were found based on faculty evaluator gender (effect size difference, 0.02 milestone levels; 95% CI for males, -0.09 to 0.11) or evaluator-evaluatee gender pairing (effect size difference, -0.02 milestone levels; 95% CI for interaction, -0.05 to 0.01).

CONCLUSIONS AND RELEVANCE:
Although male and female residents receive similar evaluations at the beginning of residency, the rate of milestone attainment throughout training was higher for male than female residents across all EM subcompetencies, leading to a gender gap in evaluations that continues until graduation. Faculty should be cognizant of possible gender bias when evaluating medical trainees.

Does One Size Fit All? Examining the Application of Neurosurgery Residency Milestones Developed in the USA to a Taiwanese Culture

BACKGROUND:
The Milestone Project was launched in 2009, charging specialties to develop specific educational accomplishments required to establish clinical competency. The milestone assessment method was first introduced to Taiwan in 2013 and prior to applying milestone assessments to our medical education system, the validity and reliability of these questionnaires needed to be evaluated.

METHOD:
Twenty neurosurgical faculty members representing 3 clinical divisions and all 4 branch institutes completed milestone questionnaires for 26 residents semiannually resulting in 435 resident assessments being collected and analyzed.

RESULTS:
Cronbach's a, KR-20, and Kendall's W were used to show acceptable reliability and validity. Rater consistencies for non-skilled parts found that rater consistency progressively improved over time. Not all raters were able to assess the residents for the skilled parts resulting in non- assessable rates ranging from 9.5% to 89.4%. For both non-skilled and skilled items, milestone level as assessed by the staff improved as the resident progressed from R3 to R6 in the residency program and showed that the milestone achievement level for an R3 was lower than that of an R6.

CONCLUSION:
Milestone assessments have high reliability and may be a helpful assessment tool. Although milestone assessment can provide thorough feedback concerning performance and the content of the training program, they may not perfectly suit all residency-training programs, especially in different countries or different cultures. Modifications should be done before applying milestones to different areas so that the results can truly reflect the progress and condition of the training and learning process.

Evaluating Surgical Residents Quickly and Easily against the Milestones Using Electronic Formative Feedback

OBJECTIVE:
This study was conducted to assess the effectiveness of a newly implemented electronic web- based review system created at our institution for evaluating resident performance relative to established milestones.

DESIGN:
Retrospective review of data collected from a survey of general surgery faculty and residents.

SETTING:
Tertiary care teaching hospital system and independent academic medical center.

PARTICIPANTS:
A total of 12 general surgery faculty and 17 general surgery residents participated in this study. The survey queried the level of satisfaction before and after the adoption of QuickNotes using several statements scored on a 5-point scale, with 1 being the lowest rating as "not satisfied," and 5 being the highest rating as "completely satisfied."

RESULTS:
The weighted average improvements from pre- to post-QuickNotes implementation for the faculty responding to the survey ranged from 10% to 40%; weighted average improvements for the residents responding to the survey ranged from 5% to 73%. For the survey of faculty, both sets of weighted averages tended to be higher than the weighted average for the resident's survey responses. The highest rated topic was the faculty's level of satisfaction with the "frequency to provide feedback" with a post-QuickNotes implementation weighted average of 4.25, closely followed by the residents' level of satisfaction with the "evaluation includes positive feedback" with a post-QuickNotes implementation weighted average of 4.24. The most notable increases in weighted averages from preimplementation to postimplementation were noted for "overall satisfaction" (20% increase for faculty, 37% for residents), "reflects actual criteria that matter" (36% increase for faculty, 73% for residents), faculty "opportunity for follow- up" (increase of 40%), resident "reflects overall trends" (increase of 37%), and resident "provides new information about my performance" (increase of 37%).

CONCLUSIONS:
Our institutional adoption of QuickNotes into the resident evaluation process has been associated with an overall increased level of satisfaction in the evaluation process by both faculty and residents. The design of QuickNotes facilitates its integration into the resident training environment, as it is web based, easy to use, and has no additional cost over the standard New Innovations subscription. Although it is designed to capture snapshots of trainee behavior and performance, monthly reports through QuickNotes can be used effectively in conjunction with the more traditional end-of-rotation evaluations to show trends, identify areas of strength that should be reinforced, demonstrate areas needing improvement, allow for a more tailored individual education plan to be developed, and permit a more accurate determination of milestone progression.

Implementation of a Needs-Based, Online Feedback Tool for Anesthesia Residents With Subsequent Mapping of the Feedback to the ACGME Milestones

BACKGROUND:
Optimizing feedback that residents receive from faculty is important for learning. The goals of this study were to (1) conduct focus groups of anesthesia residents to define what constitutes optimal feedback; (2) develop, test, and implement a web-based feedback tool; and (3) then map the contents of the written comments collected on the feedback tool to the Accreditation Council for Graduate Medical Education (ACGME) anesthesiology milestones.

METHODS:
All 72 anesthesia residents in the program were invited to participate in 1 of 5 focus groups scheduled over a 2-month period. Thirty-seven (51%) participated in the focus groups and completed a written survey on previous feedback experiences. On the basis of the focus group input, an initial online feedback tool was pilot-tested with 20 residents and 62 feedback sessions, and then a final feedback tool was deployed to the entire residency to facilitate the feedback process. The completed feedback written entries were mapped onto the 25 ACGME anesthesiology milestones.

RESULTS:
Focus groups revealed 3 major barriers to good feedback: (1) too late such as, for example, at the end of month-long clinical rotations, which was not useful because the feedback was delayed; (2) too general and not specific enough to immediately remedy behavior; and (3) too many in that the large number of evaluations that existed that were unhelpful such as those with unclear behavioral anchors compromised the overall feedback culture. Thirty residents (42% of 72 residents in the program) used the final online feedback tool with 121 feedback sessions with 61 attendings on 15 rotations at 3 hospital sites. The number of feedback tool uses per resident averaged 4.03 (standard deviation 5.08, median 2, range 1-21, 25th-75th % quartile 1-4). Feedback tool uses per faculty averaged 1.98 (standard deviation 3.2, median 1, range 1-25, 25th-75th % quartile 1-2). For the feedback question item "specific learning objective demonstrated well by the resident," this yielded 296 milestone-specific responses. The majority (71.3%) were related to the patient care competency, most commonly the anesthetic plan and conduct (35.8%) and airway management (11.1%) milestones; 10.5% were related to the interpersonal and communication skills competency, most commonly the milestones communication with other professionals (4.4%) or with patients and families (4.4%); and 8.4% were related to the practice-based learning and improvement competency, most commonly self-directed learning (6.1%). For the feedback tool item "specific learning objective that resident may improve," 67.0% were related to patient care, most commonly anesthetic plan and conduct (33.5%) followed by use/interpretation of monitoring and equipment (8.5%) and airway management (8.5%); 10.2% were related to practice-based learning and improvement, most commonly self-directed learning (6.8%); and 9.7% were related to the systems-based practice competency.

CONCLUSIONS:
Resident focus groups recommended that feedback be timely and specific and be structured around a tool. A customized online feedback tool was developed and implemented. Mapping of the free-text feedback comments may assist in assessing milestones. Use of the feedback tool was lower than expected, which may indicate that it is just 1 of many implementation steps required for behavioral and culture change to support a learning environment with frequent and useful feedback.

Number of Weeks Rotating in the Emergency Department Has a Greater Effect on Ultrasound Milestone Competency than a Dedicated Ultrasound Rotation

OBJECTIVES:
Ultrasound (US) is vital to modern emergency medicine (EM). Across residencies, there is marked variability in US training. The "goal-directed focused US" part of the Milestones Project states that trainees must correctly acquire and interpret images to achieve a level 3 milestone. Standardized methods by which programs teach these skills have not been established. Our goal was to determine whether residents could achieve level 3 with or without a dedicated US rotation.

METHODS:
Thirty-three first- and second-year residents were assigned to control (no rotation) and intervention (US rotation) groups. The intervention group underwent a 2-week curriculum in vascular access, the aorta, echocardiography, focused assessment with sonography for trauma, and pregnancy. To test acquisition, US-trained emergency medicine physicians administered an objective structured clinical examination. To test interpretation, residents had to identify normal versus abnormal findings. Mixed- model logistic regression tested the association of a US rotation while controlling for confounders: weeks in the emergency department (ED) as a resident, medical school US rotation, and postgraduate years.

RESULTS:
For image acquisition, medical school US rotation and weeks in the ED as a resident were significant ($P = .03$; $P = .04$) whereas completion of a US rotation and postgraduate years were not significant. For image interpretation, weeks in the ED as a resident was the only significant predictor of performance ($P = .002$) whereas completion of a US rotation and medical school US rotation were not significant.

CONCLUSIONS:
To achieve a level 3 milestone, weeks in the ED as a resident were significant for mastering image acquisition and interpretation. A dedicated US rotation did not have a significant effect. A medical school US rotation had a significant effect on image acquisition but not interpretation. Further studies are needed to best assess methods to meet US milestones.

How Effective are New Milestones Assessments at Demonstrating Resident Growth? 1 Year of Data

OBJECTIVE:
Assessment tools that accrue data for the Accreditation Council for Graduate Medical Education Milestones must evaluate residents across multiple dimensions, including medical knowledge, procedural skills, teaching, and professionalism. Our objectives were to: (1) develop an assessment tool to evaluate resident performance in accordance with the Milestones and (2) review trends in resident achievements during the inaugural year of Milestone implementation.

DESIGN:
A novel venue and postgraduate year (PGY) specific assessment tool was built, tested, and implemented for both operating room and labor and delivery "venues." Resident development of competence and independence was captured over time. To account for variable rotation schedules, the year was divided into thirds and compared using two-tailed Fisher's exact test.

SETTING:
Brigham and Women's and Massachusetts General Hospitals, Boston MA.

PARTICIPANTS:
Faculty evaluators and obstetrics and gynecology residents.

RESULTS:
A total of 822 assessments of 44 residents were completed between 9/2014 and 6/2015. The percentage of labor and delivery tasks completed "independently" increased monotonically across the start of all years: 8.4% for PGY-1, 60.3% for PGY-2, 73.7% for PGY-3, and 87.5% for PGY-4. Assessments of PGY-1 residents demonstrated a significant shift toward "with minimal supervision" and "independent" for the management of normal labor (p = 0.03). PGY-3 residents demonstrated an increase in "able to be primary surgeon" in the operating room, from 36% of the time in the first 2/3 of the year, to 62.3% in the last 1/3 (p< 0.01).

CONCLUSION:
Assessment tools developed to assist with Milestone assignments capture the growth of residents over time and demonstrate quantifiable differences in achievements between PGY classes. These tools will allow for targeted teaching opportunities for both individual residents and residency programs.

Piloting the Mobile Medical Milestones Application (M3App©): A Multi-Institution Evaluation

Page C, Reid A, Coe CL, Beste J, Fagan B, Steinbacher E, Newton WP. Fam Med. 2017 Jan;49 (1): 35-41.

BACKGROUND AND OBJECTIVES:
Competency-based evaluation of the Accreditation Council for Graduate Medical Education (ACGME) Milestones requires the development of new evaluation tools that can better capture learners' behavior. This study describes the implementation and initial assessment of an innovative point-of-care mobile application, the M3App©, linked to the Family Medicine Milestones.

METHODS:
Seven family medicine residency programs in North Carolina implemented the M3App.© Program faculty and residents were surveyed prior to implementation regarding current evaluation methods and their quality and use and acceptability of electronic evaluation tools. Surveys were repeated after implementation for comparison.

RESULTS:
All seven programs successfully implemented the M3App. Most faculty members found the tool well designed, easy to use, beneficial to the quality and efficiency of feedback they provide, and to their knowledge of Milestones. Residents reported significant increases in the volume and quality of written feedback they receive.

CONCLUSIONS:
The M3App provides an efficient, convenient tool for assessing Milestones that can improve the quantity and quality of feedback residents receive from faculty. Improved faculty perception of knowledge of Milestones after M3App implementation suggests that the tool is also effective for faculty development.

Connecting Milestones to Observable Clinical Performance through Standardized Pediatric Cardiology Rotation Evaluations: Report and Reflections on a Novel Pilot Project

Frank LH, Koenig PR, Srivastava S. Progress in Pediatric Cardiology. 44:11-15. doi:10.1016/j.ppedcard.2016.12.005.

ABSTRACT:
Uniform evaluations were designed and implemented by pediatric cardiology program directors. The evaluation forms incorporated ACGME/ABP subcompetencies into clinical observations. Fellow milestone assessment was achievable through the use of the evaluation forms. Early survey results suggest that these were an improvement over existing tools.

Using the ACGME Milestones for Resident Self-Evaluation and Faculty Engagement

BACKGROUND:
Since July 2014 General Surgery residency programs have been required to use the Accreditation Council for Graduate Medical Education milestones twice annually to assess the progress of their trainees. We felt this change was a great opportunity to use this new evaluation tool for resident self- assessment and to furthermore engage the faculty in the educational efforts of the program.

METHODS:
We piloted the milestones with postgraduate year (PGY) II and IV residents during the 2013/2014 academic year to get faculty and residents acquainted with the instrument. In July 2014, we implemented the same protocol for all residents. Residents meet with their advisers quarterly. Two of these meetings are used for milestones assessment. The residents perform an independent self- evaluation and the adviser grades them independently. They discuss the evaluations focusing mainly on areas of greatest disagreement. The faculty member then presents the resident to the clinical competency committee (CCC) and the committee decides on the final scores and submits them to the Accreditation Council for Graduate Medical Education website. We stored all records anonymously in a MySQL database. We used Anova with Tukey post hoc analysis to evaluate differences between groups. We used intraclass correlation coefficients and Krippendorff's a to assess interrater reliability.

RESULTS:
We analyzed evaluations for 44 residents. We created scale scores across all Likert items for each evaluation. We compared score differences by PGY level and raters (self, adviser, and CCC). We found highly significant increases of scores between most PGY levels (p < 0.05). There were no significant score differences per PGY level between the raters. The interrater reliability for the total score and 6 competency domains was very high (ICC: 0.87-0.98 and a: 0.84-0.97). Even though this milestone evaluation process added additional work for residents and faculty we had very good participation (93.9% by residents and 92.9% by faculty) and feedback was generally positive.

CONCLUSION:
Even though implementation of the milestones has added additional work for general surgery residency programs, it has also opened opportunities to furthermore engage the residents in reflection and self- evaluation and to create additional venues for faculty to get involved with the educational process within the residency program. Using the adviser as the initial rater seems to correlate closely with the final CCC assessment. Self-evaluation by the resident is a requirement by the RRC and the milestones seem to be a good instrument to use for this purpose. Our early assessment suggests the milestones provide a useful instrument to track trainee progression through theirresidency.

A Multi-Source Feedback Tool for Measuring a Subset of Pediatrics Milestones

BACKGROUND:
The Pediatrics Milestones Assessment Pilot employed a new multisource feedback (MSF) instrument to assess nine Pediatrics Milestones among interns and subinterns in the inpatient context.

OBJECTIVE:
To report validity evidence for the MSF tool for informing milestone classification decisions.

METHODS:
We obtained MSF instruments by different raters per learner per rotation. We present evidence for validity based on the unified validity framework.

RESULTS:
One hundred and ninety two interns and 41 subinterns at 18 Pediatrics residency programs received a total of 1084 MSF forms from faculty (40%), senior residents (34%), nurses (22%), and other staff (4%). Variance in ratings was associated primarily with rater (32%) and learner (22%). The milestone factor structure fit data better than simpler structures. In domains except professionalism, ratings by nurses were significantly lower than those by faculty and ratings by other staff were significantly higher. Ratings were higher when the rater observed the learner for longer periods and had a positive global opinion of the learner. Ratings of interns and subinterns did not differ, except for ratings by senior residents. MSF-based scales correlated with summative milestone scores.

CONCLUSION:
We obtain moderately reliable MSF ratings of interns and subinterns in the inpatient context to inform some milestone assignments.

Entrusting Observable Practice Activities and Milestones over the 36 Months of an Internal Medicine Residency

Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, O'Toole JK, Mathis B, Mueller C, Sall D, Tolentino J, Schauer DP. Acad Med. 2016 Oct;91(10):1398-1405.

PURPOSE:
Competency-based medical education and milestone reporting have led to increased interest in work- based assessments using entrustment over time as an assessment framework. Little is known about data collected from these assessments during residency. This study describes the results of entrustment of discrete work-based skills over 36 months in the University of Cincinnati internal medicine (IM) residency program.

METHOD:
Attending physician and peer/allied health assessors provided entrustment ratings of resident performance on work-based observable practice activities (OPAs) mapped to Accreditation Council for Graduate Medicine Education/American Board of Internal Medicine reporting milestones for IM. These data were translated into milestones data and tracked longitudinally. The authors analyzed data from this new entrustment system's first 36 months (July 2012-June 2015).

RESULTS:
During the 36-month period, assessors made 364,728 milestone assessments (mapped from OPAs) of 189 residents. Residents received an annualized average of 83 assessment encounters, producing means of 3,987 milestone assessments and 4,325 words of narrative assessment. Mean entrustment ratings (range 1-5) from all assessors for all milestones rose from 2.46 for first-month residents to 3.92 for 36th-month residents (r = 0.9252, P < .001). Attending physicians' entrustment ratings were lower than peer/allied health assessors' ratings. Medical knowledge and patient care milestones were rated lower than professionalism and interpersonal and communication skills milestones.

CONCLUSIONS:
Entrustment of milestones appears to rise progressively over time, with differences by assessor type, competency, milestone, and resident. Further research is needed to elucidate the validity of these data in promotion, remediation, and reporting decisions.

Patients, Nurses, and Physicians Working Together to Develop a Discharge Entrustable Professional Activity Assessment Tool

Meade LB, Suddarth KH, Jones RR, Zaas AK, Albanese T, Yamazaki K, O'Malley CW. Acad Med. 2016 Oct;91(10):1388-1391.

PROBLEM:
The Accreditation Council for Graduate Medical Education milestones were written by physicians and thus may not reflect all the behaviors necessary for physicians to optimize their performance as a key member of an interprofessional team.

APPROACH:
From April to May 2013, the authors, Educational Research Outcomes Collaborative leaders, assembled interprofessional team discussion groups, including patients or family members, nurses, physician trainees, physician educators, and other staff (optional), at 11 internal medicine (IM) programs. Led by the site's principal investigator, the groups generated a list of physician behaviors related to the entrustable professional activity (EPA) of a safe and effective discharge of a patient from the hospital, and prioritized those behaviors.

OUTCOMES:
A total of 182 behaviors were listed, with lists consisting of between 10 and 29 behaviors. Overall, the site principal investigators described all participants as emerging from the activity with a new understanding of the complexity of training physicians for the discharge EPA. The authors batched behaviors into six components of a safe and effective discharge: medication reconciliation, discharge summary, patient/caregiver communication, team communication, active collaboration, and anticipation of posthospital needs. Specific, high-priority behavior examples for each component were identified, and an assessment tool for direct observation was developed for the discharge EPA.

NEXT STEPS:
The authors are currently evaluating trainee and educator perceptions of the assessment tool after implementation in 15 IM programs. Additional next steps include developing tools for other EPAs, as well as a broader evaluation of patient outcomes in the era of milestone-based assessment.

Learnings from the Pilot Implementation of Mobile Medical Milestones Application

Page CP, Reid A, Coe CL, Carlough M, Rosenbaum D, Beste J, Fagan B, Steinbacher E, Jones G, Newton WP. J Grad Med Educ. 2016 Oct;8(4):569-575.

BACKGROUND:
Implementation of the educational milestones benefits from mobile technology that facilitates ready assessments in the clinical environment. We developed a point-of-care resident evaluation tool, the Mobile Medical Milestones Application (M3App), and piloted it in 8 North Carolina family medicine residency programs.

OBJECTIVE:
We sought to examine variations we found in the use of the tool across programs and explored the experiences of program directors, faculty, and residents to better understand the perceived benefits and challenges of implementing the new tool.

METHODS:
Residents and faculty completed presurveys and postsurveys about the tool and the evaluation process in their program. Program directors were interviewed individually. Interviews and open- ended survey responses were analyzed and coded using the constant comparative method, and responses were tabulated under themes.

RESULTS:
Common perceptions included increased data collection, enhanced efficiency, and increased perceived quality of the information gathered with the M3App. Residents appreciated the timely, high-quality feedback they received. Faculty reported becoming more comfortable with the tool over time, and a more favorable evaluation of the tool was associated with higher utilization. Program directors reported improvements in faculty knowledge of the milestones and resident satisfaction with feedback.

CONCLUSIONS:
Faculty and residents credited the M3App with improving the quality and efficiency of resident feedback. Residents appreciated the frequency, proximity, and specificity of feedback, and faculty reported the app improved their familiarity with the milestones. Implementation challenges included lack of a physician champion and competing demands on faculty time.

Using Milestones as Evaluation Metrics during an Emergency Medicine Clerkship

Quinn SM, Worrilow CC, Jayant DA, Bailey B, Eustice E, Kohlhepp J, Rogers R, Kane BG. J Emerg Med. 2016 Oct;51(4):426-431.

BACKGROUND:
The Accreditation Council for Graduate Medical Education's (ACGME) Milestones presumes graduating medical students will enter residency proficient at Milestone level 1 for 23 skills. The Next Accreditation System now includes Milestones for each postgraduate specialty, and it is unlikely that schools will document every emergency medicine (EM) applicant's EM-specific skills in their performance evaluation.

OBJECTIVES:
The goals of this research were to determine if assessment of the Milestones was feasible during a medical student clerkship and examine the proportion of medical students performing at Milestone level 1.

METHODS:
This study was conducted at a center with Liaison Committee on Medical Education-approved medical training and a 4-year EM residency. Using traditional clerkship, we studied the feasibility of an ACGME EM Milestones-based clerkship assessment. Data led to redesign of the clerkship and its evaluation process, including all level 1 anchor(s) to add "occasionally" (>60%), "usually" (>80%), and "always" (100%) on a Likert scale to on-shift assessment forms.

RESULTS:
During the feasibility phase (2013-14), 75 students rotated though the clerkship; 55 evaluations were issued and 50 contained the Milestone summary. Eight deficiencies were noted in Milestone 12 and three in Milestone 14. After changes, 49 students rotated under the new evaluation rubric. Of 575 completed on-shift evaluations, 16 Milestone deficiencies were noted. Of 41 institutional evaluations issued, only one student had deficiencies noted, all of which pertained to patient care. All evaluations in this second cohort contained each student's Milestone proficiency.

CONCLUSIONS:
Assessment of the Milestones is feasible. Communication of ACGME EM Milestone proficiency may identify students who require early observation or remediation. The majority of students meet the anchors for the Milestones, suggesting that clerkship assessment with the ACGME EM Milestones does not adequately differentiate students.

Shortening the Miles to the Milestones: Connecting EPA-Based Evaluations to ACGME Milestone Reports for Internal Medicine Residency Programs

ABSTRACT:
The Next Accreditation System requires internal medicine training programs to provide the Accreditation Council for Graduate Medical Education (ACGME) with semiannual information about each resident's progress in 22 subcompetency domains. Evaluation of resident "trustworthiness" in performing entrustable professional activities (EPAs) may offer a more tangible assessment construct than evaluations based on expectations of usual progression toward competence. However, translating results from EPA-based evaluations into ACGME milestone progress reports has proven to be challenging because the constructs that underlay these two systems differ. The authors describe a process to bridge the gap between rotation-specific EPA-based evaluations and ACGME milestone reporting. Developed at the University of Washington in 2012 and 2013, this method involves mapping EPA-based evaluation responses to "milestone elements," the narrative descriptions within the columns of each of the 22 internal medicine subcompetencies. As faculty members complete EPA-based evaluations, the mapped milestone elements are automatically marked as "confirmed." Programs can maintain a database that tallies the number of times each milestone element is confirmed for a resident; these data can be used to produce graphical displays of resident progress along the internal medicine milestones. Using this count of milestone elements allows programs to bridge the gap between faculty assessments of residents based on rotation-specific observed activities and semiannual ACGME reports based on the internal medicine milestones. Although potentially useful for all programs, this method is especially beneficial to large programs where clinical competency committee members may not have the opportunity for direct observation of all residents.

Mapping Direct Observations from Objective Structured Clinical Examinations to the Milestones across Specialties

BACKGROUND:
Little is known about residents' performance on the milestones at the institutional level. Our institution formed a work group to explore this using an institutional-level curriculum and residents' evaluation of the milestones.

OBJECTIVE:
We assessed whether beginner-level milestones for interpersonal and communication skills (ICS) related to observable behaviors in ICS-focused objective structured clinical examinations (OSCEs) for postgraduate year (PGY) 1 residents across specialties.

METHODS:
The work group compared ICS subcompetencies across 12 programs to identify common beginner-level physician-patient communication milestones. The selected ICS milestone sets were compared for common language with the ICS-OSCE assessment tool-the Kalamazoo Essential Elements of Communication Checklist-Adapted (KEECC-A). To assess whether OSCE scores related to ICS milestone scores, all PGY-1 residents from programs that were part of Next Accreditation System Phase 1 were identified; their OSCE scores from July 2013 to June 2014 and ICS subcompetency scores from December 2014 were compared.

RESULTS:
The milestones for 10 specialties and the transitional year had at least 1 ICS subcompetency that related to physician-patient communication. The language of the ICS beginner-level milestones appears similar to behaviors outlined in the KEECC-A. All 60 residents with complete data received at least a beginner- level ICS subcompetency score and at least a satisfactory score on all 3 OSCEs.

CONCLUSIONS:
The ICS-OSCE scores for PGY-1 residents appear to relate to beginner-level milestones for physician- patient communication across multiple specialties.

Rating the Quality of Entrustable Professional Activities: Content Validation and Associations with the Clinical Context

BACKGROUND:
Entrustable professional activities (EPAs) have been developed to assess resident physicians with respect to Accreditation Council for Graduate Medical Education (ACGME) competencies and milestones. Although the feasibility of using EPAs has been reported, we are unaware of previous validation studies on EPAs and potential associations between EPA quality scores and characteristics of educational programs.

OBJECTIVES:
Our aim was to validate an instrument for assessing the quality of EPAs for assessment of internal medicine residents, and to examine associations between EPA quality scores and features of rotations.

DESIGN:
This was a prospective content validation study to design an instrument to measure the quality of EPAs that were written for assessing internal medicine residents.

PARTICIPANTS:
Residency leadership at Mayo Clinic, Rochester participated in this study. This included the Program Director, Associate program directors and individual rotation directors.

INTERVENTIONS:
The authors reviewed salient literature. Items were developed to reflect domains of EPAs useful for assessment. The instrument underwent further testing and refinement. Each participating rotation director created EPAs that they felt would be meaningful to assess learner performance in their area. These 229 EPAs were then assessed with the QUEPA instrument to rate the quality of each EPA.

MAIN MEASURES:
Performance characteristics of the QUEPA are reported. Quality ratings of EPAs were compared to the primary ACGME competency, inpatient versus outpatient setting and specialty type.

KEY RESULTS:
QUEPA tool scores demonstrated excellent reliability (ICC range 0.72 to 0.94). Higher ratings were given to inpatient versus outpatient (3.88, 3.66; p = 0.03) focused EPAs. Medical knowledge EPAs scored significantly lower than EPAs assessing other competencies (3.34, 4.00; p < 0.0001).

CONCLUSIONS:
The QUEPA tool is supported by good validity evidence and may help in rating the quality of EPAs developed by individual programs. Programs should take care when writing EPAs for the outpatient setting or to assess medical knowledge, as these tended to be rated lower.

Reflections in a Time of Transition: Orthopaedic Faculty and Resident Understanding of Accreditation Schemes and Opinions on Surgical Skills Feedback

INTRODUCTION:
Orthopaedic surgery is one of the first seven specialties that began collecting Milestone data as part of the Accreditation Council for Graduate Medical Education's Next Accreditation System (NAS) rollout. This transition from process-based advancement to outcome-based education is an opportunity to assess resident and faculty understanding of changing paradigms, and opinions about technical skill evaluation.

METHODS:
In a large academic orthopaedic surgery residency program, residents and faculty were anonymously surveyed. A total of 31/32 (97%) residents and 29/53 (55%) faculty responded to Likert scale assessments and provided open-ended responses. An internal end-of-rotation audit was conducted to assess timeliness of evaluations. A mixed-method analysis was utilized, with nonparametric statistical testing and a constant-comparative qualitative method.

RESULTS:
There was greater familiarity with the six core competencies than with Milestones or the NAS ($p<0.05$). A majority of faculty and residents felt that end-of-rotation evaluations were not adequate for surgical skills feedback. Fifty-eight per cent of residents reported that end-of- rotation evaluations were rarely or never filled out in a timely fashion. An internal audit demonstrated that more than 30% of evaluations were completed over a month after rotation end. Qualitative analysis included themes of resident desire for more face-to-face feedback on technical skills after operative cases, and several barriers to more frequent feedback.

DISCUSSION:
The NAS and outcome-based education have arrived. Residents and faculty need to be educated on this changing paradigm. This transition period is also a window of opportunity to address methods of evaluation and feedback. In our orthopaedic residency, trainees were significantly less satisfied than faculty with the amount of technical and surgical skills feedback being provided to trainees. The quantitative and qualitative analyses converge on one theme: a desire for frequent, explicit, timely feedback after operative cases. To overcome the time-limited clinical environment, feedback tools need to be easily integrated and efficient. Creative solutions may be needed to truly achieve outcome-based graduate medical education.

Entrustable Professional Activities: Ten Things Radiologists Do

RATIONALE AND OBJECTIVES:
Learner assessment in medical education has undergone tremendous change over the past two decades. During this time frame, the concept of Entrustable Professional Activities (EPAs) was introduced to guide the faculty when making competency-based decisions on the level of supervision required by trainees. EPAs are gaining momentum in medical education as a basis for decisions related to transitioning from residency training to clinical practice. The purpose of this article is twofold: (1) define EPAs for radiology (EPA-R) and (2) illustrate radiology-specific examples of these EPAs.

MATERIALS AND METHODS:
A multi-institutional work group composed of members of the Alliance of Directors and Vice Chairs of Education in Radiology convened at the 2015 Association of University Radiologists annual meeting to discuss radiology EPAs. The EPAs initially developed by the Accreditation Council for Graduate Medical Education (ACGME) Radiology milestone work group and the resultant ACGME Radiology milestones formed the basis for this discussion.

RESULTS:
A total of 10 radiology EPAs and illustrative vignettes were developed to help radiology educators and trainees better understand milestone assessment and how this translates to the necessary skills and responsibilities of practicing radiologists. Examples of EPA mapping to the ACGME subcompetencies and methods of assessment were included.

CONCLUSIONS:
EPAs offer an opportunity to improve our approach to training by increasing our focus on how we provide appropriate supervision to our residents and assess their progress. In this work, through suggested lists and vignettes, we have attempted to establish the framework for further discussion and development of EPA-Rs.

Comprehensive Observations of Resident Evolution: A Novel Method for Assessing Procedure- Based Residency Training

BACKGROUND:
Assessment of surgical skills in the operating room remains a challenge. Increasing documentation requirements of the Accreditation Council for Graduate Medical Education are necessitating mechanisms to document trainee competence without hindering operative turnover. The authors created a comprehensive electronic resource to facilitate plastic surgery training program compliance with changes mandated by Next Accreditation System Milestones and the ACGME.

METHODS:
In 2013, the authors implemented the Comprehensive Observations of Resident Evolution, or CORE, a Web-based tool to assess plastic surgery residents. It comprises a rapid electronic assessment of resident operating room performance completed after each surgery; a data dashboard displaying graphical summaries of resident progress by case, Milestone, or current procedural terminology code; and an electronic Milestones tracker (MileMarker), which enables ongoing trainee assessments.

RESULTS:
From January through October of 2014, 24 residents completed nearly 1300 Operative Entrustability Assessments. Thirty-eight percent of residents reported more immediate feedback regarding operative performance. The assessment demonstrates construct validity, which distinguishes novice residents from experienced residents. Individual case data identify resident-specific operative strengths and weaknesses. Using assessment data, the first two Clinical Competency Committee reviews were 81 percent and 87 percent shorter than Milestones pilot test site reports (average, 11.5 and 8 minutes versus 60 minutes per resident, respectively).

CONCLUSIONS:
Comprehensive Observations of Resident Evolution is capable of capturing operative performance data on all operating room cases by primary current procedural terminology code. It increases immediate attending/trainee feedback and assessment transparency, enables trainee self-monitoring, and informs end-of-rotation reviews, program-wide assessments, and tailoring of training to address specific needs. It is a valuable resource for tracking resident progress in real-time while maintaining compliance with evolving ACGME requirements.

Driving Care Quality: Aligning Trainee Assessment and Supervision through Practical Application of Entrustable Professional Activities, Competencies, and Milestones

ABSTRACT:
To address the long-standing challenge of meaningful trainee assessment, the authors reviewed and expanded on the Accountable Assessment for Quality Care and Supervision (AAQCS) equation. The equation proposes that care quality is the product of the interaction between trainee performance (measured by workplace assessment) and supervision (required level of intervention to ensure care quality) in the context of the environment where the care occurs: Trainee performance × Appropriate supervision = Safe, effective patient-centered care. Assessing trainee performance and matching that performance to "appropriate" supervision, however, is fraught with challenges. The authors suggest a unifying framework that integrates entrustable professional activities (EPAs), competencies, and milestones to inform trainee assessment and supervision, thereby enabling the practical application of the AAQCS equation in the workplace. Because the unit of measure for an EPA is the outcome of whether the trainee can safely and effectively perform the professional activity without supervision, the proposed unifying framework directly aligns with the dependent variable in the AAQCS equation: care quality. The value of applying a unifying framework that integrates EPAs, competencies, and milestones to the AAQCS equation in the clinical learning environment lies in its ability to provide supervisors with a shared mental model of performance expectations for trainees, reducing unwanted variability and improving assessment accuracy; guidance for aligning performance milestones of trainees with the needed level of supervisor intervention to ensure care quality; and substrate for specific feedback to improve the trainee's professional development as a way to ensure future care quality.

A Pilot Study of Orthopaedic Resident Self-Assessment Using a Milestones' Survey Just Prior to Milestones Implementation

OBJECTIVE:
To pilot test if Orthopaedic Surgery residents could self-assess their performance using newly created milestones, as defined by the Accreditation Council on Graduate Medical Education.

METHODS:
In June 2012, an email was sent to Program Directors and administrative coordinators of the 154 accredited Orthopaedic Surgery Programs, asking them to send their residents a link to an online survey. The survey was adapted from the Orthopaedic Surgery Milestone Project. Completed surveys were aggregated in an anonymous, confidential database. SAS 9.3 was used to perform the analyses.

RESULTS:
Responses from 71 residents were analyzed. First and second year residents indicated through self- assessment that they had substantially achieved Level 1 and Level 2 milestones. Third year residents reported they had substantially achieved 30/41, and fourth year residents, all Level 3 milestones. Fifth year, graduating residents, reported they had substantially achieved 17 Level 4 milestones, and were extremely close on another 15. No milestone was rated at Level 5, the maximum possible. Earlier in training, Patient Care and Medical Knowledge milestones were rated lower than the milestones reflecting the other four competencies of Practice Based Learning and Improvement, Systems Based Practice, Professionalism, and Interpersonal Communication. The gap was closed by the fourth year.

CONCLUSIONS:
Residents were able to successfully self-assess using the 41 Orthopaedic Surgery milestones. Respondents' rate improved proficiency over time. Graduating residents report they have substantially, or close to substantially, achieved all Level 4 milestones. Milestone self- assessment may be a useful tool as one component of a program's overall performance assessment strategy.

Competency Evaluations in the Next Accreditation System: Contributing to Guidelines and Implications

CONSTRUCT:
This study examines validity evidence of end-of-rotation evaluation scores used to measure competencies and milestones as part of the Next Accreditation System (NAS) of the Accreditation Council for Graduate Medical Education (ACGME).

BACKGROUND:
Since the implementation of the milestones, end-of-rotation evaluations have surfaced as a potentially useful assessment method. However, validity evidence on the use of rotation evaluation scores as part of the NAS has not been studied. This article examines validity evidence for end-of-rotation evaluations that can contribute to developing guidelines that support the NAS.

APPROACH:
Data from 2,701 end-of-rotation evaluations measuring 21 out of 22 Internal Medicine milestones for 142 residents were analyzed (July 2013-June 2014). Descriptive statistics were used to measure the distribution of ratings by evaluators (faculty, n = 116; fellows, n = 59; peer- residents, n = 131), by postgraduate years. Generalizability analysis and higher order confirmatory factor analysis were used to examine the internal structure of ratings. Psychometric implications for combining evaluation scores using composite score reliability were examined.

RESULTS:
Milestone ratings were significantly higher for each subsequent year of training (15/21 milestones). Faculty evaluators had greater variability in ratings across milestones, compared to fellows and residents; faculty ratings were generally correlated with milestone ratings from fellows (r = .45) and residents (r = .25), but lower correlations were found for Professionalism and Interpersonal and Communication Skills. The ¢-coefficient was .71, indicating good reliability. Internal structure supported a 6-factor solution, corresponding to the hierarchical relationship between the milestones and the 6 core competencies. Evaluation scores corresponding to Patient Care, Medical Knowledge, and Practice- Based Learning and Improvement had higher correlations to milestones reported to the ACGME. Mean evaluation ratings predicted problem residents (odds ratio = 5.82, p < .001).

CONCLUSIONS:
Guidelines for rotation evaluations proposed in this study provide useful solutions that can help program directors make decisions on resident progress and contribute to assessment systems in graduate medical education.

Use of a Formal Assessment Instrument for Evaluation of Resident Operative Skills in Pediatric Neurosurgery

OBJECTIVE:
Currently there is no standardized tool for assessment of neurosurgical resident performance in the operating room. In light of enhanced requirements issued by the Accreditation Council for Graduate Medical Education's Milestone Project and the Matrix Curriculum Project from the Society of Neurological Surgeons, the implementation of such a tool seems essential for objective evaluation of resident competence. Beyond compliance with governing body guidelines, objective assessment tools may be useful to direct early intervention for trainees performing below the level of their peers so that they may be given more hands-on teaching, while strong residents can be encouraged by faculty members to progress to conducting operations more independently with passive supervision. The aims of this study were to implement a validated assessment tool for evaluation of operative skills in pediatric neurosurgery and determine its feasibility and reliability.

METHODS:
All neurosurgery residents completing their pediatric rotation over a 6-month period from January 1, 2014, to June 30, 2014, at the authors' institution were enrolled in this study. For each procedure, residents were evaluated by means of a form, with one copy being completed by the resident and a separate copy being completed by the attending surgeon. The evaluation form was based on the validated Objective Structured Assessment of Technical Skills for Surgery (OSATS) and used a 5-point Likert-type scale with 7 categories: respect for tissue; time and motion; instrument handling; knowledge of instruments; flow of operation; use of assistants; and knowledge of specific procedure. Data were then stratified by faculty versus resident (self-) assessment; postgraduate year level; and difficulty of procedure. Descriptive statistics (means and SDs) were calculated, and the results were compared using the Wilcoxon signed-rank test and Student t- test. A p value < 0.05 was considered statistically significant.

RESULTS:
Six faculty members, 1 fellow, and 8 residents completed evaluations for 299 procedures, including 32 ventriculoperitoneal (VP) shunt revisions, 23 VP shunt placements, 19 endoscopic third ventriculostomies, and 18 craniotomies for tumor resection. There was no significant difference between faculty and resident self-assessment scores overall or in any of the 7 domains scores for each of the involved residents. On self-assessment, senior residents scored themselves significantly higher (p < 0.02) than junior residents overall and in all domains except for "time and motion." Faculty members scored senior residents significantly higher than junior residents only for the "knowledge of instruments" domain (p = 0.05). When procedure difficulty was considered, senior residents' scores from faculty members were significantly higher (p = 0.04) than the scores given to junior residents for expert procedures only. Senior residents' self- evaluation scores were significantly higher than those of junior residents for both expert (p = 0.03) and novice (p = 0.006) procedures.

CONCLUSIONS:
OSATS is a feasible and reliable assessment tool for the comprehensive evaluation of neurosurgery resident performance in the operating room. The authors plan to use this tool to assess resident operative skill development and to improve direct resident feedback.

Assessing EM Patient Safety and Quality Improvement Milestones Using a Novel Debate Format

ABSTRACT
Graduate medical education is increasingly focused on patient safety and quality improvement; training programs must adapt their curriculum to address these changes. We propose a novel curriculum for emergency medicine (EM) residency training programs specifically addressing patient safety, systems- based management, and practice-based performance improvement, called "EM Debates." Following implementation of this educational curriculum, we performed a cross-sectional study to evaluate the curriculum through resident self-assessment. Additionally, a cross-sectional study to determine the ED clinical competency committee's (CCC) ability to assess residents on specific competencies was performed. Residents were overall very positive towards the implementation of the debates. Of those participating in a debate, 71% felt that it improved their individual performance within a specific topic, and 100% of those that led a debate felt that they could propose an evidence-based approach to a specific topic. The CCC found that it was easier to assess milestones in patient safety, systems-based management, and practice-based performance improvement (sub-competencies 16, 17, and 19) compared to prior to the implementation of the debates. The debates have been a helpful venue to teach EM residents about patient safety concepts, identifying medical errors, and process improvement.

Gearing Up for Milestones in Surgery: Will Simulation Play a Role?

BACKGROUND:
The Consortium of American College of Surgeons-Accredited Education Institutes was created to promote patient safety through the use of simulation, develop new education and technologies, identify best practices, and encourage research and collaboration.

METHODS:
During the 7th Annual Meeting of the Consortium, leaders from a variety of specialties discussed how simulation is playing a role in the assessment of resident performance within the context of the Milestones of the Accreditation Council for Graduate Medical Education as part of the Next Accreditation System.

CONCLUSION:
This report presents experiences from several viewpoints and supports the utility of simulation for this purpose.

Direct Observation Assessment of Milestones: Problems with Reliability

INTRODUCTION:
Emergency medicine (EM) milestones are used to assess residents' progress. While some milestone validity evidence exists, there is a lack of standardized tools available to reliably assess residents.

Inherent to this is a concern that we may not be truly measuring what we intend to assess. The purpose of this study was to design a direct observation milestone assessment instrument supported by validity and reliability evidence. In addition, such a tool would further lend validity evidence to the EM milestones by demonstrating their accurate measurement.

METHODS:
This was a multi-center, prospective, observational validity study conducted at eight institutions. The Critical Care Direct Observation Tool (CDOT) was created to assess EM residents during resuscitations. This tool was designed using a modified Delphi method focused on content, response process, and internal structure validity. Paying special attention to content validity, the CDOT was developed by an expert panel, maintaining the use of the EM milestone wording. We built response process and internal consistency by piloting and revising the instrument. Raters were faculty who routinely assess residents on the milestones. A brief training video on utilization of the instrument was completed by all. Raters used the CDOT to assess simulated videos of three residents at different stages of training in a critical care scenario. We measured reliability using Fleiss' kappa and interclass correlations.

RESULTS:
Two versions of the CDOT were used: one used the milestone levels as global rating scales with anchors, and the second reflected a current trend of a checklist response system. Although the raters who used the CDOT routinely rate residents in their practice, they did not score the residents' performances in the videos comparably, which led to poor reliability. The Fleiss' kappa of each of the items measured on both versions of the CDOT was near zero.

CONCLUSION:
The validity and reliability of the current EM milestone assessment tools have yet to be determined. This study is a rigorous attempt to collect validity evidence in the development of a direct observation assessment instrument. However, despite strict attention to validity evidence, inter-rater reliability was low. The potential sources of reducible variance include rater- and instrument-based error. Based on this study, there may be concerns for the reliability of other EM milestone assessment tools that are currently in use.

Pathology Milestones: Assessing Clinical Competency by Committee

ABSTRACT:
All Accreditation Council for Graduate Medical Education accredited pathology residency training programs are now required to evaluate residents using the new Pathology Milestones assessment tool. Similar to implementation of the 6 Accreditation Council for Graduate Medical Education competencies a decade ago, there have been challenges in implementation of the new milestones for many residency programs. The pathology department at the University of Iowa has implemented a process that divides the labor of the task in rating residents while also maintaining consistency in the process. The process is described in detail, and some initial trends in milestone evaluation are described and discussed. Our experience indicates that thoughtful implementation of the Pathology Milestones can provide programs with valuable information that can inform curricular changes.

Navigating the Next Accreditation System: A Dashboard for the Milestones

INTRODUCTION:
In July 2014, all residency programs accredited by the Accreditation Council for Graduate Medical Education (ACGME) were enrolled in a new system called the Next Accreditation System. Residency programs may not be clear on how best to comply with these new accreditation requirements. Large amounts of data must be collected, evaluated, and submitted twice a year to the council's Web-based data collection system. One challenge is that the new "end-of-rotation" evaluations must reflect specialty-specific milestones, on which many faculty members are not well versed. Like other residency programs, we tried to address the challenges using our local resources.

METHODS:
We used our existing electronic goals and objectives for each rotation coupled with appropriate end-of- rotation evaluations reflecting the specialty-specific milestones through a process of editing and mapping.

RESULTS:
Data extracted from these evaluations were added to an interactive dashboard that also contained evaluations on additional program-specific modifiers of residents' performance. A resident's final overall performance was visually represented on a plot graph. The novel dashboard included features to save evaluations for future comparisons and to track residents' progress during their entire training. It proved simple to use and was able to reduce the time needed for each resident evaluation to 5 to 10 minutes.

CONCLUSION:
This tool has made it much easier and less challenging for the members of our Clinical Competency Committee to start deliberation about each resident's performance.

Interprofessional Collaboration Milestones: Advocating for Common Assessment Criteria in Graduate Medical Education

BACKGROUND:
Milestone-based assessments of resident physicians inform critical decisions regarding resident competence and advancement. Thus, it is essential that milestone evaluations are based upon strong validity evidence and that consistent evaluation criteria are used across residency programs. A common approach to assessment of interprofessional collaboration milestones is particularly important since standardized measures of individual resident competence in interprofessional collaboration have not been established.

DISCUSSION:
We propose that assessments of interprofessional collaboration in graduate medical education meet common criteria, namely, these assessments should: 1) measure competency of an individual resident, 2) occur in the context of an interprofessional team, 3) be ascertained via direct observation of the resident, 4) be performed in a real-world clinical practice setting (such as a hospital ward, outpatient clinic, or operating room). We present the evidence-based rationale for these criteria and cite examples of published assessment instruments that fulfill one or more of the criteria, however further research is needed to ensure fidelity of assessments. The proposed criteria may assist residency educators as they endeavor to provide robust and consistent assessments of interprofessional collaboration milestones.

Development and Validation of an Assessment of Regional Anesthesia Ultrasound Interpretation Skills

Woodworth GE, Carney PA, Cohen JM, Kopp SL, Vokach-Brodsky LE, Horn JL, Missair A, Banks SE, Dieckmann NF, Maniker RB. Reg Anesth Pain Med. 2015 Jul-Aug;40(4):306-14. doi: 10.1097/AAP.0000000000000236.

BACKGROUND:
Interpretation of ultrasound images and knowledge of anatomy are essential skills for ultrasound-guided peripheral nerve blocks. Competency-based educational models promoted by the Accreditation Council for Graduate Medical Education require the development of assessment tools for the achievement of different competency milestones to demonstrate the longitudinal development of skills that occur during training.

METHODS:
A rigorous study guided by psychometric principles was undertaken to identify and validate the domains and items in an assessment of ultrasound interpretation skills for regional anesthesia. A survey of residents, academic faculty, and community anesthesiologists, as well as video recordings of experts teaching ultrasound-guided peripheral nerve blocks, was used to develop short video clips with accompanying multiple choice-style questions. Four rounds of pilot testing produced a 50-question assessment that was subsequently administered online to residents, fellows, and faculty from multiple institutions.

RESULTS:
Test results from 90 participants were analyzed with Item Response Theory model fitting indicating that a 47-item subset of the test fits the model well (P = 0.11). There was a significant linear relation between expected and predicted item difficulty (P < 0.001). Overall test scores increased linearly with higher levels of formal anesthesia training, regional anesthesia training, number of ultrasound-guided blocks performed per year, and a self-rating of regional anesthesia skill (all P < 0.001).

CONCLUSIONS:
This study provides evidence for the reliability, content validity, and construct validity of a 47- item multiple choice-style online test of ultrasound interpretation skills for regional anesthesia, which can be used as an assessment of competency milestone achievement in anesthesiology training.

Developing a Comprehensive Resident Education Evaluation System in the Era of Milestone Assessment

OBJECTIVES:
In an effort to move training programs toward competency-based education, the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which organizes specific milestones regarding resident skills, knowledge, and abilities along a continuum. In order to foster innovation and creativity, the ACGME has provided programs with minimal guidelines regarding the optimal way to approach these milestones.

METHODS:
The education team at UT Southwestern embraced the milestones and developed a process in which performance assessment methods were critically evaluated, mapped onto an extrapolated performance list corresponding to the areas required by the ACGME milestones, and filled gaps in the previous system by modifying evaluation tools and creating new program components.

RESULTS:
Although the authors are early in the evolution of applying the new milestones system, this approach has thus far allowed them to comprehensively evaluate the residents and the program in an efficient and effective fashion, with notable improvements compared to the prior approach.

CONCLUSIONS:
The authors hope that these experiences can inform others embarking upon similar journeys with the milestones.

Implementation of Nephrology Subspecialty Curricular Milestones

ABSTRACT:
Beginning in the 2014-2015 training year, the US Accreditation Council for Graduate Medical Education (ACGME) required that nephrology Clinical Competency Committees assess fellows' progress toward 23 subcompetency "context nonspecific" internal medicine subspecialty milestones. Fellows' advancement toward the "ready for unsupervised practice" target milestone now is tracked in each of the 6 competencies: Patient Care, Medical Knowledge, Professionalism, Interpersonal Communication Skills, Practice-Based Learning and Improvement, and Systems-Based Practice. Nephrology program directors and subspecialty societies must define nephrology-specific "curricular milestones," mapped to the nonspecific ACGME milestones. Although the ACGME goal is to produce data that can discriminate between successful and underperforming training programs, the approach is at risk to produce biased, inaccurate, and unhelpful information. We map the ACGME internal medicine subspecialty milestones to our previously published nephrology-specific milestone schema and describe entrustable professional activities and other objective assessment tools that inform milestone decisions. Mapping our schema onto the ACGME subspecialty milestone reporting form allows comparison with the ACGME subspecialty milestones and the curricular milestones developed by the American Society of Nephrology Program Directors. Clinical Competency Committees may easily adapt and directly translate milestone decisions reached using our schema onto the ACGME internal medicine subspecialty competency milestone-reporting format.

Piloting a Structured Practice Audit to Assess ACGME Milestones in Written Handoff Communication in Internal Medicine

BACKGROUND:
Written communication skills are integral to patient care handoffs. Residency programs require feasible assessment tools that provide timely formative and summative feedback, ideally linked to the Accreditation Council for Graduate Medical Education Milestones.

OBJECTIVE:
We describe the use of 1 such tool-UPDATED-to assess written handoff communication skills in internal medicine interns.

METHODS:
During 2012-2013, the authors piloted a structured practice audit at 1 academic institution to audit written sign-outs completed by 45 interns, using the UPDATED tool, which scores 7 aspects of sign-out communication linked to milestones. Intern sign-outs were audited by trained faculty members throughout the year. Results were incorporated into intern performance reviews and Clinical Competency Committees.

RESULTS:
A total of 136 sign-outs were audited (averaging 3.1 audits per intern). In the first trimester, 14 interns (31%) had satisfactory audit results. Five interns (11%) had critical deficiencies and received immediate feedback, and the remaining 26 (58%) were assigned future audits due to missing audits or unsatisfactory scores. In the second trimester, 21 interns (68%) had satisfactory results, 1 had critical deficiencies, and 9 (29%) required future audits. Nine of the 10 remaining interns in the final trimester had satisfactory audits. Faculty time was estimated at 10 to 15 minutes per sign-out audited.

CONCLUSIONS:
The UPDATED audit is a milestone-based tool that can be used to assess written sign-out communication skills in internal medicine residency programs. Future work is planned to adapt the tool for use by senior supervisory residents to appraise sign-outs in real time.

Use of Emergency Medicine Milestones as Items on End-of-Shift Evaluations Results in Overestimates of Residents' Proficiency Level

BACKGROUND:
The emergency medicine milestones were developed to provide more objective resident assessment than current methods. However, little is known about the best practices for applying the milestones in resident assessment.

OBJECTIVE:
We examined the utility of end-of-shift evaluations (ESEs) constructed using the milestones in resident assessment.

METHODS:
We developed 14 daily ESEs, each of which included 9 or 10 emergency medicine milestones. Postgraduate year (PGY)-1 and PGY-2 residents were assessed on milestone levels 1 through 3; PGY-3 and PGY-4 residents were assessed on levels 3 through 5. Each milestone was rated on a nominal scale (yes, no, or not applicable). The Clinical Competency Committee combined the ESE data with data from other assessments to determine each resident's proficiency level for the emergency medicine subcompetencies. We used descriptive statistics to summarize resident ESEs and milestone levels. We analyzed differences in ESE score across PGY levels using t tests and analyses of variance.

RESULTS:
Faculty completed 763 ESEs on 33 residents with a range of 2 to 54 (median=22) ESEs per resident. Faculty rarely (8%, 372 of 4633) rated a resident as not achieving a milestone on the ESEs. Analyses of variance revealed that ESE scores on level 3 milestones did not differ significantly by PGY level. There was poor agreement between ESE scores and Clinical Competency Committee ratings.

CONCLUSIONS:
The ESEs constructed using the milestones resulted in grade or milestone inflation. Our results do not support using milestones as a stand-alone assessment tool.

Clinical Assessment and Management Examination—Outpatient (CAMEO): Its Validity and Use in a Surgical Milestones Paradigm

OBJECTIVES:
Clinical Assessment and Management Examination--Outpatient (CAMEO) is a metric for evaluating the clinical performance of surgery residents. The aim of this study was to investigate the measurement characteristics of CAMEO and propose how it might be used as an evaluation tool within the general surgery milestones project.

DESIGN:
A total of 117 CAMEO evaluations were gathered and used for analysis. Internal consistency reliability was estimated, and item characteristics were explored. A Kruskal-Wallis procedure was performed to discern how well the instrument discriminated between training levels. An exploratory factor analysis was also conducted to understand the dimensionality of the evaluation.

SETTING:
CAMEO evaluations were collected from 2 departments of surgery geographically located in the Midwestern United States. Combined, the participating academic institutions graduate approximately 18 general surgery residents per year.

PARTICIPANTS:
In this retrospective data analysis, the number of evaluations per resident ranged from 1 to 7, and evaluations were collected from 2006 to 2013. For the purpose of data analysis, residents were classified as interns (postgraduate year 1 [PGY1]), juniors (PGY2-3), or seniors (PGY4-5).

RESULTS:
CAMEO scores were found to have high internal consistency (Cronbach's a = 0.96), and all items were highly correlated (≥ 0.86) to composite CAMEO scores. Scores discriminated between senior residents (PGY4-5) and lower level residents (PGY1-3). Per an exploratory factor analysis, CAMEO was revealed to measure a single dimension of "clinical competence."

CONCLUSIONS:
The findings of this research aligned with related literature and verified that CAMEO scores have desirable measurement properties, making CAMEO an attractive resource for evaluating the clinical performance of surgery residents.

A Faculty Development Program to Reduce Rater Error on Milestone-Based Assessments

Raj JM, Thorn, PM. Journal of Graduate Medical Education. December 2014. doi: http://dx.doi.org/10.4300/JGME-D-14-00161.1

BACKGROUND:
Rater errors, such as halo/reverse halo, range restriction, and leniency errors, are frequently cited as threats to the validity of resident assessment by faculty.

OBJECTIVE:
We studied whether participation in faculty development on the use of a new Milestone-based assessment tool reduced rater error for participants compared to individuals who did not participate.

METHODS:
We reviewed evaluations of resident Milestones completed by faculty at the end of rotations between July 2012 and June 2013. The 2 Milestones in each competency with the greatest number of ratings were selected for analysis.

RESULTS:
A total of 412 evaluations were analyzed, including 217 completed by faculty who participated in the development activity, and 240 completed by nonparticipant faculty. All evaluations that contained identical scores for all Milestones (16%) were completed by nonparticipant faculty (x2 5 37.498, P, .001). Faculty who had participated in development assigned a wider range of scores and lower minimum scores to residents, and provided the highest ratings for residents less frequently (P, .001) than nonparticipants.

CONCLUSIONS:
Faculty who participated in education about the Milestones demonstrated significantly less halo, range restriction, and leniency errors than faculty members who did not participate. These findings support a recommendation to develop a cadre of ''core faculty'' by training them in the use of Milestone assessment tools, and making them responsible for a significant portion of resident assessments.

Immediate Surgical Skills Feedback in the Operating Room Using "SurF" Cards

BACKGROUND:
Ensuring residents develop operative skills requires application of the principles of guided learning, deliberate practice, and directed feedback.

OBJECTIVE:
We sought to create and implement a tool to promote procedural "key" step review and immediate feedback on surgical skills, and examined faculty and resident satisfaction with surgical skills feedback.

METHODS:
We created surgical skills feedback (SurF) cards for 8 gynecologic procedures. Faculty/fellows and residents completed prestudy surveys querying frequency of preoperative key step review and satisfaction with surgical skill feedback, a SurF card each time 1 of 8 procedures was performed, and poststudy surveys to evaluate for changes.

RESULTS:
Prestudy surveys were completed by 31 faculty/fellows and 20 residents, with 55% (17 of 31) of the faculty/fellows and 5% (1 of 20) of the residents reporting key step review before surgery. All reported low satisfaction rates with feedback frequency, quality, and timeliness. After implementation of SurF cards, preoperative key step review occurred in 78% (82 of 105) of the procedures. Twenty-one faculty/fellows (68%) and 16 residents (80%) completed our poststudy survey. Faculty/fellows reported statistically similar key step review (n = 15 [71%], P = .23), while residents reported that key step review had significantly improved (n = 6 [38%], P = .01). Resident satisfaction with feedback frequency (5% to 50%, P = .002) and quality (15% to 50%, P = .02) increased significantly.

CONCLUSIONS:
The SurF cards we developed facilitated procedural key step review, were associated with significantly improved resident satisfaction with surgical feedback, and could prove helpful with outcomes assessments, such as Accreditation Council for Graduate Medical Education-required documentation of Milestone attainment.

Assessment of Resident Operative Performance Using a Real-Time Mobile Web System: Preparing for the Milestone Age

OBJECTIVE:
To satisfy trainees' operative competency requirements while improving feedback validity and timeliness using a mobile Web-based platform.

DESIGN:
The Southern Illinois University Operative Performance Rating Scale (OPRS) was embedded into a website formatted for mobile devices. From March 2013 to February 2014, faculty members were instructed to complete the OPRS form while providing verbal feedback to the operating resident at the conclusion of each procedure. Submitted data were compiled automatically within a secure Web-based spreadsheet. Conventional end-of-rotation performance (CERP) evaluations filed 2006 to 2013 and OPRS performance scores were compared by year of training using serial and independent-samples t tests. The mean CERP scores and OPRS overall resident operative performance scores were directly compared using a linear regression model. OPRS mobile site analytics were reviewed using a Web- based reporting program.

SETTING:
Large university-based general surgery residency program.

PARTICIPANTS:
General Surgery faculty used the mobile Web OPRS system to rate resident performance. Residents and the program director reviewed evaluations semiannually.

RESULTS:
Over the study period, 18 faculty members and 37 residents logged 176 operations using the mobile OPRS system. There were 334 total OPRS website visits. Median time to complete an evaluation was 45 minutes from the end of the operation, and faculty spent an average of 134 seconds on the site to enter 1 assessment. In the 38,506 CERP evaluations reviewed, mean performance scores showed a positive linear trend of 2% change per year of training (p = 0.001). OPRS overall resident operative performance scores showed a significant linear (p = 0.001), quadratic (p = 0.001), and cubic (p = 0.003) trend of change per year of clinical training, reflecting the resident operative experience in our training program. Differences between postgraduate year-1 and postgraduate year-5 overall performance scores were greater with the OPRS (mean = 0.96, CI: 0.55-1.38) than with CERP measures (mean = 0.37, CI: 0.34-0.41). Additionally, there were consistent increases in each of the OPRS subcategories.

CONCLUSIONS:
In contrast to CERPs, the OPRS fully satisfies the Accreditation Council for Graduate Medical Education and American Board of Surgery operative assessment requirements. The mobile Web platform provides a convenient interface, broad accessibility, automatic data compilation, and compatibility with common database and statistical software. Our mobile OPRS system encourages candid feedback dialog and generates a comprehensive review of individual and group-wide operative proficiency in real time.

Entrustable Professional Activities: Making Sense of the Emergency Medicine Milestones

BACKGROUND:
The Next Accreditation System (NAS) is being implemented by the Accreditation Council for Graduate Medical Education with seven specialties, including Emergency Medicine (EM), which began in July 2013. The NAS represents a more structured method of accreditation, with dependence on outcomes and less emphasis on educational process. A key component of the NAS is the individual resident semiannual reporting of the Milestone proficiency levels for all sub-competencies, which are more specific areas of domain for the general competencies. All specialties are struggling to some extent with developing assessment mechanisms for the Milestones. At the heart of this struggle is the conceptualization of the Milestones themselves- descriptors of the individual. In practice, faculty assess clinical care provided to the patient by the resident. This creates difficulty for faculty to assign a resident to a specific sub-competency proficiency level when their focus has been on assessment of clinical care.

OBJECTIVES:
The objectives of this article include the discussion of whether Entrustable Professional Activities (EPAs) could be defined and linked to milestones in a way that, once implemented, could inform Clinical Competency Committees of the Milestone proficiency reporting.

DISCUSSION:
EPAs are units of professional work, or clinical care that may help translate aspects of clinical care into Milestone proficiencies. This article explores EPAs in depth, and discusses how EPAs may be used within EM as one method of assigning proficiency levels to residents.

CONCLUSIONS:
EPAs may be a useful tool to inform Milestone proficiency placement of residents. Because EPAs are based on clinical descriptions rather than individual physician descriptions, there may be less faculty development needed for Milestone sub-competency assessment.

Entrustment and Mapping of Observable Practice Activities for Resident Assessment

Warm EJ, Mathis BR, Held JD, Pai S, Tolentino J, Ashbrook L, Lee CK, Lee D, Wood S, Fichtenbaum CJ, Schauer D, Munyon R, Mueller C. J Gen Intern Med. 2014 Aug;29(8):1177-82. doi: 10.1007/s11606-014-2801-5. Epub 2014 Feb 21.

ABSTRACT:
Entrustable Professional Activities (EPAs) and the Next Accreditation System reporting milestones reduce general competencies into smaller evaluable parts. However, some EPAs and reporting milestones may be too broad to use as direct assessment tools. We describe our internal medicine residency curriculum and assessment system, which uses entrustment and mapping of observable practice activities (OPAs) for resident assessment. We created discrete OPAs for each resident rotation and learning experience. In combination, these serve as curricular foundation and tools for assessment. OPA performance is measured via a 5-point entrustment scale, and mapped to milestones and EPAs. Entrustment ratings of OPAs provide an opportunity for immediate structured feedback of specific clinical skills, and mapping OPAs to milestones and EPAs can be used for longitudinal assessment, promotion decisions, and reporting. Direct assessment and demonstration of progressive entrustment of trainee skill over time are important goals for all training programs. Systems that use OPAs mapped to milestones and EPAs provide the opportunity for achieving both, but require validation.

Assessing Competency in Practice-Based Learning: A Foundation for Milestones in Learning Portfolio Entries

BACKGROUND:
Graduate medical education is undergoing a dramatic shift toward competency-based assessment of learners. Competency assessment requires clear definitions of competency and validated assessment methods. The purpose of this study is to identify criteria used by surgical educators to judge competence in Practice-Based Learning and Improvement (PBL&I) as demonstrated in learning portfolios.

METHODS:
A total of 6 surgical learning and instructional portfolio entries served as documents to be assessed by 3 senior surgical educators. These faculty members were asked to rate and then identify criteria used to assess PBL&I competency. Individual interviews and group discussions were conducted, recorded, and transcribed to serve as the study dataset. Analysis was performed using qualitative methodology to identify themes for the purpose of defining competence in PBL&I. The assessment themes derived are presented with narrative examples to describe the progression of competency.

RESULTS:
The collaborative coding process resulted in identification of 7 themes associated with competency in PBL&I related to surgical learning and instructional portfolio entries: (1) self- awareness regarding effect of actions; (2) identification and thorough description of learning goals; (3) cases used as catalyst for reflection; (4) reconceptualization with appropriate use and critique of cited literature; (5) communication skills/completeness of entry template; (6) description of future behavioral change; and (7) engagement in process--identifies as personally relevant.

CONCLUSIONS:
The identified themes are consistent with and complement other criteria emerging from reflective practice literature and experiential learning theory. This study provides a foundation for further development of a tool for assessing learner portfolios consistent with the Accreditation Council for Graduate Medical Education's Next Accreditation System requirements.

Programmatic Assessment of Level 1 Milestones in Incoming Interns

OBJECTIVES:
With the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System, emergency medicine (EM) residency programs will be required to report residents' progress through the EM milestones. The milestones include five progressively advancing skill levels, with Level 1 defining the skill set of a medical school graduate and Level 5, that of an attending physician. The ACGME stresses that multiple forms of assessment should be used to ensure capture of the multifaceted competencies. The objective of this study was to determine the feasibility and results of programmatic assessment of Level 1 milestones using multisource assessments for incoming EM interns in July.

METHODS:
The study population was interns starting in 2012 and 2013. Interns' Level 1 milestone assessment was done with four distinct methods: 1) the postgraduate orientation assessment (POA) by the Graduate Medical Education Office for all incoming interns (this multistation examination covers nine of the EM milestones and includes standardized patient cases, task completion, and computer-based stations); 2) direct observation of patient encounters by core faculty using a milestones-based clinical skills competency checklist; 3) the global monthly assessment at the end of the intern orientation month that was updated to reflect the EM milestones; and 4) faculty assessment during procedural labs. These occurred during the July orientation month that included the POA, clinical shifts, didactic sessions, and procedure labs.

RESULTS:
In the POA, interns were competent in 48% to 93% of the milestones assessed. Overall, competency was 70% to 80%, with low scores noted in aseptic technique (patient care Milestone 13 [PC13]) and written and verbal hand-off (interpersonal communications skills [ICS] 2). In overall communication, 70% of interns demonstrated competency. In excess of 80% demonstrated competency in critical values interpretation (PC3), informed consent (PC9), pain assessment (PC11), and geriatric functional assessment (PC3). On direct observation, almost all Level 1 milestones were achieved (93% to 100%); however, only 78% of interns achieved competency in pharmacotherapy (PC5). On global monthly evaluations, all interns met Level 1 milestones.

CONCLUSIONS:
A multisource assessment of EM milestones is feasible and useful to determine Level 1 milestones achievement for incoming interns. A structured assessment program, used in conjunction with more traditional forms of evaluation such as global monthly evaluations and direct observation, is useful for identifying deficits in new trainees and may be able inform the creation of early intervention programs.

Complexity in Graduate Medical Education: A Collaborative Education Agenda for Internal Medicine and Geriatric Medicine

ABSTRACT:
Internal medicine residents today face significant challenges in caring for an increasingly complex patient population within ever-changing education and health care environments. As a result, medical educators, health care system leaders, payers, and patients are demanding change and accountability in graduate medical education (GME). A 2012 Society of General Internal Medicine (SGIM) retreat identified medical education as an area for collaboration between internal medicine and geriatric medicine. The authors first determined a short-term research agenda for resident education by mapping selected internal medicine reporting milestones to geriatrics competencies, and listing available sample learner assessment tools. Next, the authors proposed a strategy for long-term collaboration in three priority areas in clinical medicine that are challenging for residents today: (1) team-based care, (2) transitions and readmissions, and (3) multi-morbidity. The short-term agenda focuses on learner assessment, while the long-term agenda allows for program evaluation and improvement. This model of collaboration in medical education combines the resources and expertise of internal medicine and geriatric medicine educators with the goal of increasing innovation and improving outcomes in GME targeting the needs of our residents and their patients.

Putting the Pediatrics Milestones into Practice: A Consensus Roadmap and Resource Analysis

ABSTRACT:
The Accreditation Council for Graduate Medical Education has partnered with member boards of the American Board of Medical Specialties to initiate the next steps in advancing competency-based assessment in residency programs. This initiative, known as the Milestone Project, is a paradigm shift from traditional assessment efforts and requires all pediatrics residency programs to report individual resident progression along a series of 4 to 5 developmental levels of performance, or milestones, for individual competencies every 6 months beginning in June 2014. The effort required to successfully make this shift is tremendous given the number of training programs, training institutions, and trainees. However, it holds great promise for achieving training outcomes that align with patient needs; developing a valid, reliable, and meaningful way to track residents' development; and providing trainees with a roadmap for learning. Recognizing the resources needed to implement this new system, the authors, all residency program leaders, provide their consensus view of the components necessary for implementing and sustaining this effort, including resource estimates for completing this work. The authors have identified 4 domains: (1) Program Review and Development of Stakeholders and Participants, (2) Assessment Methods and Validation, (3) Data and Assessment System Development, and (4) Summative Assessment and Feedback.

This work can serve as a starting point and framework for collaboration with program, department, and institutional leaders to identify and garner necessary resources and plan for local and national efforts that will ensure successful transition to milestones-based assessment.

Integrating the NAS Milestones and Handheld Technology to Improve Residency Training and Assessment

OBJECTIVE:
To incorporate the use of an intuitive and robust assessment tool in conjunction with the Next Accreditation System Milestones to maximize opportunities for trainee performance feedback and continuous trainee assessment, with the long-term goal of increasing the rate of performance improvement and mastery of knowledge and surgical skills.

DESIGN:
Pilot study.

SETTING:
Johns Hopkins Medicine, Baltimore, MD. Primary, tertiary, and quaternary clinical care; institutional environment.

PARTICIPANTS:
Experimental group: two randomly selected postgraduate year-1 integrated training program residents per year for 2 consecutive years from the Department of Plastic and Reconstructive Surgery.

CONTROL GROUP:
Traditionally trained residents from the integrated training program in the Department of Plastic and Reconstructive Surgery. Study duration: 7 years (until residents complete residency training).

ANTICIPATED RESULTS:
This assessment strategy would create large amounts of informative data on trainees, which can be cross-referenced to determine trainee progress. Assessment data would be collected continuously from all faculty surgeons. Comparisons of faculty and resident self-assessments would facilitate resident evaluations. Ease of use of the data collection structure would improve faculty evaluation compliance and timely resident case report completion.

CONCLUSIONS:
Improving the efficiency and efficacy of competency documentation is critical. Using portable technologies is an intuitive way to improve the trainee assessment process. We anticipate that this 2- pronged approach to trainee assessments would quickly provide large amounts of informative data to better assess trainee progress and inform Milestone assessments in a manner that facilitates immediate feedback. Assessments of faculty and resident satisfaction would help us further refine the assessment process as needed. If successful, this format could easily be implemented by other training programs.

Initial Performance of a Modified Milestones Global Evaluation Tool for Semiannual Evaluation of Residents by Faculty

Borman KR, Augustine R, Leibrandt T, Pezzi CM, Kukora JS. J Surg Educ. 2013 Nov- Dec;70(6):739-49. doi: 10.1016/j.jsurg.2013.08.004.

OBJECTIVES:
To determine whether faculty could successfully evaluate residents using a competency-based modified Milestones global evaluation tool.

DESIGN:
A program's leadership team modified a draft Surgery Milestones Working Group summative global assessment instrument into a modified Milestones tool (MMT) for local use during faculty meetings devoted to semiannual resident review. Residents were scored on 15 items spanning all competencies using an 8-point graphic response scale; unstructured comments also were solicited. Arithmetic means were computed at the resident and postgraduate year cohort levels for items and competency item sets. Score ranges (highest minus lowest score) were calculated; variability was termed "low" (range <2.0 points), "moderate" (range = 2.0), or "high" (range >2.0). A subset of "low" was designated "small" (1.0- 1.9). Trends were sought among item, competency, and total Milestones scores. MMT correlations with examination scores and multisource (360°) assessments were explored. The success of implementing MMT was judged using published criteria for educational assessment methods.

SETTING:
Fully accredited, independently sponsored residency.

PARTICIPANTS:
Program leaders and 22 faculty members (71% voluntary, mean 12y of experience).

RESULTS:
Twenty-six residents were assessed, yielding 7 to 13 evaluations for MMT per categorical resident and 3 to 6 per preliminary trainee. Scores spanned the entire response scale. All MMT evaluations included narrative comments. Individual resident score variability was low (96% within competencies and 92% across competencies). Subset analysis showed that small variations were common (35% within competencies and 54% across competencies). Postgraduate year cohort variability was higher (61% moderate or high within competencies and 50% across competencies). Cohort scores at the item, competency, and total score levels exhibited rising trajectories, suggesting MMT construct validity. MMT scores did not demonstrate concurrent validity, correlating poorly with other metrics. The MMT met multiple criteria for good assessment.

CONCLUSIONS:
A modified Milestones global evaluation tool can be successfully adopted for semiannual assessments of resident performance by volunteer faculty members.

Early Feedback on the Use of the Internal Medicine Reporting Milestones in Assessment of Resident Performance

Aagaard E, Kane GC, Conforti L, Hood S, Caverzagie KJ, Smith C, Chick DA, Holmboe ES, Iobst WF. J Grad Med Educ. 2013 Sep;5(3):433-8

BACKGROUND:
The educational milestones were designed as a criterion-based framework for assessing resident progression on the 6 Accreditation Council for Graduate Medical Education competencies.

OBJECTIVE:
We obtained feedback on, and assessed the construct validity and perceived feasibility and utility of, draft Internal Medicine Milestones for Patient Care and Systems-Based Practice.

METHODS:
All participants in our mixed-methods study were members of competency committees in internal medicine residency programs. An initial survey assessed participant and program demographics; focus groups obtained feedback on the draft milestones and explored their perceived utility in resident assessment, and an exit survey elicited input on the value of the draft milestones in resident assessment. Surveys were tabulated using descriptive statistics. Conventional content analysis method was used to assess the focus group data.

RESULTS:
Thirty-four participants from 17 programs completed surveys and participated in 1 of 6 focus groups. Overall, the milestones were perceived as useful in formative and summative assessment of residents. Participants raised concerns about the length and complexity of some draft milestones and suggested specific changes. The focus groups also identified a need for faculty development. In the exit survey, most participants agreed that the Patient Care and Systems-Based Practice Milestones would help competency committees assess trainee progress toward independent practice.

CONCLUSIONS:
Draft reporting milestones for 2 competencies demonstrated significant construct validity in both the content and response process and the perceived utility for the assessment of resident performance. To ensure success, additional feedback from the internal medicine community and faculty development will be necessary.

Playing with Curricular Milestones in the Educational Sandbox: Q-Sort Results from an Internal Medicine Educational Collaborative

PURPOSE:
In competency-based medical education, the focus of assessment is on learner demonstration of predefined outcomes or competencies. One strategy being used in internal medicine (IM) is applying curricular milestones to assessment and reporting milestones to competence determination. The authors report a practical method for identifying sets of curricular milestones for assessment of a landmark, or a point where a resident can be entrusted with increased responsibility.

METHOD:
Thirteen IM residency programs joined in an educational collaborative to apply curricular milestones to training. The authors developed a game using Q-sort methodology to identify high-priority milestones for the landmark "Ready for indirect supervision in essential ambulatory care" (EsAMB). During May to December 2010, the programs'ambulatory faculty participated in the Q-sort game to prioritize 22 milestones for EsAMB. The authors analyzed the data to identify the top 8 milestones.

RESULTS:
In total, 149 faculty units (1-4 faculty each) participated. There was strong agreement on the top eight milestones; six had more than 92% agreement across programs, and five had 75% agreement across all faculty units. During the Q-sort game, faculty engaged in dynamic discussion about milestones and expressed interest in applying the game to other milestones and educational settings.

CONCLUSIONS:
The Q-sort game enabled diverse programs to prioritize curricular milestones with interprogram and interparticipant consistency. A Q-sort exercise is an engaging and playful way to address milestones in medical education and may provide a practical first step toward using milestones in the real-world educational setting.