Milestones
Bibliography
June 2018

Sonia Sangha, MPH
Stanley J. Hamstra, PhD
# Table of Contents

- **Introduction** ................................................................................................................................. 3
- **Use of Bibliography** .......................................................................................................................... 4
- **National-Level Studies**
  - (Quantitative) ................................................................................................................................. 5
  - (Qualitative) ........................................................................................................................................ 18
- **Multi-Institutional Studies**
  - (Quantitative) ................................................................................................................................. 23
  - (Qualitative) ........................................................................................................................................ 40
- **Single Institution Studies**
  - (Quantitative) ....................................................................................................................................... 45
  - (Qualitative) ........................................................................................................................................ 75
- **Medical School Milestones Studies**
  - (Quantitative) ....................................................................................................................................... 85
  - (Qualitative) ........................................................................................................................................ 87
- **International Studies**
  - (Quantitative) ....................................................................................................................................... 89
  - (Qualitative) ........................................................................................................................................ 90
- **Commentaries/Editorials/Opinion Pieces** ...................................................................................... 91
- **Appendix**
  - Papers Describing the Initial Development of Milestones Language by Specialty
    (JGME suppl. issues March 2013 and March) .................................................................................... 120
Introduction

This bibliography provides a list of peer-reviewed published articles, as well as commentaries, editorials, and opinion pieces that specifically address the topic of the Milestones.

This list is designed to be useful in tracking validity evidence for the Milestones, as well as in accessing links to educational theory and strategies for implementation of this framework in individual residency and fellowship programs.

Additionally, the list was envisioned to serve as a springboard for further research into the validity and use of the Milestones to help develop the best evidence for decisions around resident and fellows progression within accredited programs and beyond.

This list will be updated approximately every six months.

Stanley J. Hamstra, PhD
Vice President, Milestone Research and Evaluation
shamstra@acgme.org

Sonia Sangha, MPH
Research Association for Milestones Evaluation
ssangha@acgme.org
### Use of Bibliography

The bibliography is organized according to the categories below, and is presented by year of publication (in descending order) within each category.

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Studies:</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>13</td>
</tr>
<tr>
<td>Qualitative</td>
<td>5</td>
</tr>
<tr>
<td><strong>Multi-Institution Studies:</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>17</td>
</tr>
<tr>
<td>Qualitative</td>
<td>5</td>
</tr>
<tr>
<td><strong>Single Institution Studies:</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>30</td>
</tr>
<tr>
<td>Qualitative</td>
<td>10</td>
</tr>
<tr>
<td><strong>Medical School Milestone Studies:</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>2</td>
</tr>
<tr>
<td>Qualitative</td>
<td>2</td>
</tr>
<tr>
<td><strong>International Studies:</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>1</td>
</tr>
<tr>
<td>Qualitative</td>
<td>1</td>
</tr>
<tr>
<td><strong>Commentaries/Editorials/Opinion Pieces</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>115</td>
</tr>
</tbody>
</table>

The breakdown by Year of Publication is as follows:

- 2018: 23
- 2017: 38
- 2016: 19
- 2015: 21
- 2014: 6
- 2013: 6
- 2012: 2

**Keywords Used and Search Criteria:**
- (milestone*[Title/Abstract]) AND (resident* OR fellow* OR (graduate AND medical AND education))
- Date Range: 2012-2018
- Database(s) used: EBSCO, PubMed

*Note: since the content that follows comes from a variety of sources, there may be significant variations in style and spelling with respect to terminology. For official ACGME terminology and terminology usage, refer to the ACGME Glossary of Terms.*
National Internal Medicine Milestone Ratings: Validity Evidence From Longitudinal Three-Year Follow-Up.


PURPOSE:
To evaluate validity evidence for internal medicine milestone ratings across programs for three resident cohorts by quantifying "not assessable" ratings; reporting mean longitudinal milestone ratings for individual residents; and correlating medical knowledge ratings across training years with certification examination scores to determine predictive validity of milestone ratings for certification outcomes.

METHOD:
This retrospective study examined milestone ratings for postgraduate year (PGY) 1-3 residents in United States internal medicine residency programs. Data sources included milestone ratings, program characteristics, and certification examination scores.

RESULTS:
Among 35,217 participants, there was a decreased percentage with "not assessable" ratings across years: 1,566 (22.5%) PGY1s in 2013-2014 versus 1,219 (16.6%) in 2015-2016 (P = .01), and 342 (5.1%) of PGY3s in 2013-2014 versus 177 (2.6%) in 2015-2016 (P = .04). For individual residents with three years of ratings, mean milestone ratings increased from around 3 (behaviors of an early learner or advancing resident) in PGY1 (ranging from a mean of 2.73 to 3.19 across subcompetencies) to around 4 (ready for unsupervised practice) in PGY3 (mean of 4.00 to 4.22 across subcompetencies, P < .001 for all subcompetencies). For each increase of 0.5 units in two medical knowledge (MK1, MK2) subcompetency ratings, the difference in examination scores for PGY3s was 19.5 points for MK1 (P < .001) and 19.0 for MK2 (P < .001).

CONCLUSIONS:
These findings provide evidence of validity of the milestones by showing how training programs have applied them over time and how milestones predict other training outcomes.
Straight Line Scoring by Clinical Competency Committees Using Emergency Medicine Milestones.


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:
Theorem 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.
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.
Examining the Functioning and Reliability of the Family Medicine Milestones.

BACKGROUND:
The Family Medicine (FM) Milestones are a framework designed to assess development of residents in key dimensions of physician competency. Residency programs use the milestones in semiannual reviews of resident performance from entry toward graduation.

OBJECTIVE:
To examine the functioning and reliability of the FM Milestones and to determine whether they measure the amount of a latent trait (e.g., knowledge or ability) possessed by a resident or simply indicate where a resident falls along the training sequence.

METHODS:
This study utilized the Rasch Partial Credit model to examine academic year 2014-2015 ratings for 10,563 residents from 476 residency programs (postgraduate year [PGY] 1 = 3,639; PGY-2 = 3,562; PGY-3 = 3,351; PGY-4 = 11).

RESULTS:
Reliability was exceptionally high at 0.99. Mean scores were 3.2 (SD = 1.3) for PGY-1; 5.0 (SD = 1.3) for PGY-2; 6.7 (SD = 1.2) for PGY-3; and 7.4 (SD = 1.0) for PGY-4. Keyform analysis showed a rating on 1 item was likely to be similar for all other items.

CONCLUSIONS:
Our findings suggest that FM Milestones seem to largely function as intended. Lack of spread in item difficulty and lack of variation in category probabilities show that FM Milestones do not measure the amount of a latent trait possessed by a resident, but rather describe where a resident falls along the training sequence. High reliability indicates residents are being rated in a stable manner as they progress through residency, and individual residents deviating from this rating structure warrant consideration by program leaders.
Program Director Perceptions of the General Surgery Milestones Project.


OBJECTIVE:
As a result of the Milestones Project, all Accreditation Council for Graduate Medical Education accredited training programs now use an evaluation framework based on outcomes in 6 core competencies. Despite their widespread use, the Milestones have not been broadly evaluated. This study sought to examine program director (PD) perceptions of the Milestones Project.

DESIGN, SETTING, AND PARTICIPANTS:
A national survey of general surgery PDs distributed between January and March of 2016.

RESULTS:
A total of 132 surgical PDs responded to the survey (60% response rate). Positive perceptions included value for education (55%) and evaluation of resident performance (58%), as well as ability of Milestones to provide unbiased feedback (55%) and to identify areas of resident deficiency (58%). Meanwhile, time input and the ability of Milestones to discriminate underperforming programs were less likely to be rated positively (25% and 21%, respectively). Half of PDs felt that the Milestones were an improvement over their previous evaluation system (55%).

CONCLUSIONS:
Using the Milestones as competency-based, developmental outcomes measures, surgical PDs reported perceived benefits for education and objectivity in the evaluation of resident performance. The overall response to the Milestones was generally favorable, and most PDs would not return to their previous evaluation systems. To improve future iterations of the Milestones, many PDs expressed a desire for customization of the Milestones' content and structure to allow for programmatic differences.


IMPORTANCE: US internal medicine residency programs are now required to rate residents using milestones. Evidence of validity of milestone ratings is needed.

OBJECTIVE: To compare ratings of internal medicine residents using the pre-2015 resident annual evaluation summary (RAES), a nondevelopmental rating scale, with developmental milestone ratings.

DESIGN, SETTING, AND PARTICIPANTS: Cross-sectional study of US internal medicine residency programs in the 2013-2014 academic year, including 21284 internal medicine residents (7048 postgraduate-year 1 [PGY-1], 7233 PGY-2, and 7003 PGY-3).

EXPOSURES: Program director ratings on the RAES and milestone ratings.

MAIN OUTCOMES AND MEASURES: Correlations of RAES and milestone ratings by training year; correlations of medical knowledge ratings with American Board of Internal Medicine (ABIM) certification examination scores; rating of unprofessional behavior using the 2 systems.

RESULTS: Corresponding RAES ratings and milestone ratings showed progressively higher correlations across training years, ranging among competencies from 0.31 (95% CI, 0.29 to 0.33) to 0.35 (95% CI, 0.33 to 0.37) for PGY-1 residents to 0.43 (95% CI, 0.41 to 0.45) to 0.52 (95% CI, 0.50 to 0.54) for PGY-3 residents (all P values <.05). Linear regression showed ratings differed more between PGY-1 and PGY-3 years using milestone ratings than the RAES (all P values <.001). Of the 6260 residents who attempted the certification examination, the 618 who failed had lower ratings using both systems for medical knowledge than did those who passed (RAES difference, -0.9; 95% CI, -1.0 to -0.8; P < .001; milestone medical knowledge 1 difference, -0.3; 95% CI, -0.3 to -0.3; P < .001; and medical knowledge 2 difference, -0.2; 95% CI, -0.3 to -0.2; P < .001). Of the 26 PGY-3 residents with milestone ratings indicating deficiencies on either of the 2 medical knowledge subcompetencies, 12 failed the certification examination. Correlation of RAES ratings for professionalism with residents' lowest professionalism milestone ratings was 0.44 (95% CI, 0.43 to 0.45; P < .001).

CONCLUSIONS AND RELEVANCE: Among US internal medicine residents in the 2013-2014 academic year, milestone-based ratings correlated with RAES ratings but with a greater difference across training years. Both rating systems for medical knowledge correlated with ABIM certification examination scores. Milestone ratings may better detect problems with professionalism. These preliminary findings may inform establishment of the validity of milestone-based assessment.
BACKGROUND:
High-quality assessment of resident performance is needed to guide individual residents' development and ensure their preparedness to provide patient care. To facilitate this aim, reporting milestones are now required across all internal medicine (IM) residency programs.

OBJECTIVE:
To describe initial milestone ratings for the population of IM residents by IM residency programs.

DESIGN:
Cross-sectional study.

SETTING:
IM residency programs.

PARTICIPANTS:
All IM residents whose residency program directors submitted milestone data at the end of the 2013-2014 academic year.

MEASUREMENTS:
Ratings addressed 6 competencies and 22 subcompetencies. A rating of "not assessable" indicated insufficient information to evaluate the given subcompetency. Descriptive statistics were calculated to describe ratings across competencies and training years.

RESULTS:
Data were available for all 21,774 U.S. IM residents from all 383 programs. Overall, 2,889 residents (1,621 in postgraduate year 1 [PGY-1], 902 in PGY-2, and 366 in PGY-3) had at least 1 subcompetency rated as not assessable. Summaries of average ratings by competency and training year showed higher ratings for PGY-3 residents in all competencies. Overall ratings for each of the 6 individual competencies showed that fewer than 1% of third-year residents were rated as "unsatisfactory" or "conditional on improvement." However, when subcompetency milestone ratings were used, 861 residents (12.8%) who successfully completed training had at least 1 competency with all corresponding subcompetencies graded below the threshold of "readiness for unsupervised practice."

LIMITATION: Data were derived from a point in time in the first reporting period in which milestones were used.

CONCLUSION: The initial milestone-based evaluations of IM residents nationally suggest that documenting developmental progression of competency is possible over training years. Subcompetencies may identify areas in which residents might benefit from additional feedback and experience. Future work is needed to explore how milestones are used to support residents' development and enhance residency curricula.
A Survey of Ultrasound Milestone Incorporation Into Emergency Medicine Training Programs.

OBJECTIVES:
With the introduction of the Emergency Medicine Milestone Project in 2013, residencies now assess emergency ultrasound (US) skills at regular intervals. However, it is unclear how programs are implementing the emergency US milestones and assessing competency. With the use of the milestone tool, a survey was distributed to emergency US educators to determine when programs are providing emergency US education, when residents are expected to attain competency, and whether the milestones reflect their expectations of trainees.

METHODS:
We conducted a prospective cross-sectional survey study distributed electronically to designated emergency US experts at 169 programs. Participants were queried on education and competency evaluation within the context of the milestones by designating a postgraduate year when the 5 milestone levels were taught and competency was expected. Survey findings were reported as percentages of total respondents from descriptive statistics.

RESULTS:
Responses were received from 53% of programs, and 99% were familiar with the milestones. Most programs provide level 1 (88%) and 2 (85%) instruction during postgraduate year 1. Most programs expect level 1 competency before residency (61%) and expect mastery of level 2 by the end of postgraduate year 1 (60%). Sixty-two percent believe the milestones do not accurately reflect their expectations, citing insufficient minimum scan numbers, lack of specificity, and unattainable level 5 requirements.

CONCLUSIONS:
There is substantial variability in the frequency and methods of competency evaluation using the emergency US milestones. However, most responders agree that residents should obtain level 2 competency by postgraduate year 1. Variation exists regarding what year and what skills define level 3 or greater competency.
"What Program Directors Think" III: Results of the 2014/2015 Annual Surveys of the Association of Program Directors in Radiology (APDR).


RATIONALE AND OBJECTIVES:
The Association of Program Directors in Radiology regularly surveys its members regarding issues of importance to support radiology residency programs and their directors.

MATERIALS AND METHODS:
This is an observational cross-sectional study using two Web-based surveys posed to the Association of Program Directors in Radiology membership in the fall of 2014 (49 items) and the spring of 2015 (46 items) on the subjects of importance to the members, including the Accreditation Council on Graduate Medical Education Milestones, the Non-Interpretative Skills Curriculum, the American Board of Radiology Core Examination, the effect of the new resident testing and program accreditation paradigms on training outcomes, the 2015 Residency Match, the Interventional Radiology/Diagnostic Radiology (IR/DR) Residency, and Program Director (PD)/Program Coordinator resources.

RESULTS:
Responses were collected electronically, results were tallied using SurveyMonkey software, and qualitative responses were tabulated or summarized as comments. Findings were reported during the 63rd annual meeting of the Association of University Radiologists. The maximal response rate was 33% in the fall of 2014 and 36% in the spring of 2015.

CONCLUSIONS:
PDs believed that the radiology Milestones, now largely implemented, did not affect overall resident evaluation, was not reflective of resident experience, and actually made evaluation of residents more difficult. PDs also felt that although the American Board of Radiology oral examination had been a better test for clinical practice preparedness, their new residents knew at least as much as before. There was little evidence of recall reemergence. The radiology training community saw a drop in residency applicant quality as demonstrated by the United States Medical Licensing Examination scores and clinical rotation grades. Because the new IR/DR Residency positions were to be funded at the expense of the traditional DR positions, the majority of PDs expected a negative effect of the impending IR/DR match on their DR recruitment. PDs were in favor of a unified clinical radiology curriculum similar to the Radiological Society of North America online physics modules.
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 Φ-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.
Initial Validity Analysis of the Emergency Medicine Milestones.

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) Milestones describe behavioral markers for the progressive acquisition of competencies during residency. As a key component of the Next Accreditation System, all residents are evaluated for the acquisition of specialty-specific Milestones. The objective was to determine the validity and reliability of the emergency medicine (EM) Milestones.

METHODS:
The ACGME and the American Board of Emergency Medicine performed this single-event observational study. The data included the initial EM Milestones performance ratings of all categorical EM residents submitted to the ACGME from October 31, 2013, to January 6, 2014. Mean performance ratings were determined for all 23 subcompetencies for every year of residency training. The internal consistency (reliability) of the Milestones was determined using a standardized Cronbach’s alpha coefficient.

Exploratory factor analysis was conducted to determine how the subcompetencies were interrelated.

RESULTS:
EM Milestone performance ratings were obtained on 100% of EM residents (n = 5,805) from 162 residency programs. The mean performance ratings of the aggregate and individual subcompetency scores showed discrimination between residency years, and the factor structure further supported the validity of the EM Milestones. The reliability was $\alpha = 0.96$ within each year of training.

CONCLUSIONS:
The EM Milestones demonstrated validity and reliability as an assessment instrument for competency acquisition. EM residents can be assured that this evaluation process has demonstrated validity and reliability; faculty can be confident that the Milestones are psychometrically sound; and stakeholders can know that the Milestones are a nationally standardized, objective measure of specialty-specific competency acquisition.
The emergency medicine milestones: a validation study.

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties sought to define milestones for skill and knowledge acquisition during residency training. Milestones are significant objective observable events. The milestones are listed within a structure that is derived from the ACGME general competencies. Major groups of milestones are called "subcompetencies." The original 24 subcompetencies containing 255 milestones for emergency medicine (EM) were developed through a multiorganizational group representing most EM stakeholder groups. To assure that the milestones reflected EM resident progress throughout training, the EM Milestones Working Group (EM MWG) sought to validate the individual milestones.

METHODS:
A computer-based survey was sent to all EM residency programs. The survey period began on April 30, 2012, and concluded on May 15, 2012. Respondents were asked to assign each milestone to a specific level of skill or knowledge acquisition. These levels ranged from a beginning resident to an accomplished clinician. There were two different forms that divided the milestones into two groups of 12 subcompetencies each. Surveys were randomly assigned to programs.

RESULTS:
There were five respondents (the program director and four key faculty) requested from each of the 159 residences. There were responses from 96 programs (60.4%). Of the 795 survey recipients, 28 were excluded due to prior exposure to the EM milestones. Of the remaining 767 potential respondents, 281 completed the survey (36.6%) within a 16-day period. Based on the survey results, the working group adjusted the milestones in the following ways: one entire subcompetency (teaching) was eliminated, six new milestones were created, 34 milestones were eliminated, 26 milestones were reassigned to a lower level score, and 20 were reassigned to a higher level. Nineteen milestones were edited to provide greater clarity. The final result was 227 discrete milestones among 23 subcompetencies.

CONCLUSIONS:
The EM milestones were validated through a milestone assignment process using a computer-based survey completed by program directors and key faculty. Milestones were revised in accordance with the results to better align assignment within each performance level.
National-Level Studies: Qualitative

Pediatric Program Director Minimum Milestone Expectations before Allowing Supervision of Others and Unsupervised Practice.


BACKGROUND:
The Accreditation Council for Graduate Medical Education requires semiannual Milestone reporting on all residents. Milestone expectations of performance are unknown.

OBJECTIVE:
Determine pediatric program director (PD) minimum Milestone expectations for residents prior to being ready to supervise and prior to being ready to graduate.

METHODS:
Mixed methods survey of pediatric PDs on their programs' Milestone expectations before residents are ready to supervise and before they are ready to graduate, and in what ways PDs use Milestones to make supervision and graduation decisions. If programs had no established Milestone expectations, PDs indicated expectations they considered for use in their program. Mean minimum Milestone level expectations adjusted for program size, region, and clustering of Milestone expectations by program were calculated for prior to supervise and prior to graduate. Free-text questions were analyzed using thematic analysis.

RESULTS:
The response rate was 56.8% (113/199). Most programs had no required minimum Milestone level before residents are ready to supervise (80%; 76/95) or ready to graduate (84%; 80/95). For readiness to supervise, minimum Milestone expectations PDs considered establishing for their program were highest for humanism (2.46, 95% CI: 2.21-2.71) and professionalization (2.37, 2.15-2.60). Minimum Milestone expectations for graduates were highest for help-seeking (3.14, 2.83-3.46). Main themes included the use of Milestones in combination with other information to assess learner performance and Milestones are not equally weighted when making advancement decisions.

CONCLUSIONS:
Most PDs have not established program minimum Milestones, but would vary such expectations by competency.
Competency Crosswalk: A Multispecialty Review of the Accreditation Council for Graduate Medical Education Milestones Across Four Competency Domains.


PURPOSE:
To identify common and overlapping themes among the interpersonal and communication skills (ICS), practice-based learning and improvement (PBLI), professionalism (PROF), and systems-based practice (SBP) milestones of the transitional year and 26 specialties.

METHOD:
In May 2017, milestones were accessed from the Accreditation Council for Graduate Medical Education specialties website. A thematic analysis of the ICS, PBLI, PROF, and SBP milestones was performed to determine unique and common themes across these competencies and across specialties. Keywords from the common program requirements were initially applied as codes to the milestones. Codes were then grouped into common themes.

RESULTS:
Twenty-two themes were identified: 15 (68%) were unique to a given competency (3 related to ICS, 4 related to PBLI, 5 related to PROF, and 3 related to SBP), and 7 (32%) appeared in the milestones of more than one core competency. Eleven themes (50%) were used by 20 or more specialties, and 6 themes (27%) by 10 or fewer specialties. No theme was present across all specialties.

CONCLUSIONS:
The ICS, PBLI, PROF, and SBP milestones contain multiple themes with areas of overlap among these four competencies and substantial variability across specialties. This variability may create differential expectations of residents across specialties, complicate faculty development, and make sharing assessment tools difficult. The thematic analysis provides important insights into how individual specialties interpret and operationalize the ICS, PBLI, PROF, and SBP competency domains and can inform future revisions of milestones to enable harmonization and shared understanding of these competencies across specialties where appropriate.


PURPOSE:
Quality improvement (QI) and patient safety (PS) are broadly relevant to the practice of medicine, but specialty-specific milestones demonstrate variable expectations for trainee competency in QI/PS. The purpose of this study was to develop a unifying portrait of QI/PS expectations for graduating residents irrespective of specialty.

METHOD:
Milestones from 26 residency programs representing the 24 member boards of the American Board of Medical specialties were downloaded from the Accreditation Council for Graduate Medical Education (ACGME) website in 2015. A codebook was generated by in-depth reading of all milestone sets by two authors. Using a content analytic approach, milestones were then coded by a single author, with a 10% sample double-coded by another author. Descriptive statistics were used to characterize frequency counts.

RESULTS:
Of 612 total milestones, 249 (40.7%) made mention of QI/PS. A median 10 milestones per specialty (interquartile range, 5.25-11.75) mentioned QI/PS. There were 446 individual references to QI, 423 references to PS, and another 1,065 references to QI/PS-related concepts, including patient-centered care, cost-effective practice, documentation, equity, handoffs and care transitions, and teamwork. QI/PS references reflected expectations about both individual-level practice (531/869, 61.1%) and practice within a healthcare system (338/869, 38.9%). QI and PS references were linked to all six ACGME core competencies.

CONCLUSIONS:
Although there is variability in the emphasis placed on QI/PS across specialties, overall, QI/PS is reflected in more than 40% of residency milestones. Graduating residents in all specialties are expected to demonstrate competence in QI, PS, and multiple related concepts.
Aligning In-Service Training Examinations in Plastic Surgery and Orthopaedic Surgery With Competency-Based Education.


BACKGROUND:
In-service training examinations (ITEs) are used to assess residents across specialties. However, it is not clear how they are integrated with the Accreditation Council for Graduate Medical Education Milestones and competencies.

OBJECTIVE:
This study explored the distribution of specialty-specific milestones and competencies in ITEs for plastic surgery and orthopaedic surgery.

METHODS:
In-service training examinations were publicly available for plastic surgery (PSITE) and orthopaedics (OITE). Questions on the PSITE for 2014-2016 and the OITE for 2013-2015 were mapped to the specialty-specific milestones and the 6 competencies.

RESULTS:
There was an uneven distribution of milestones and competencies in ITE questions. Nine of the 36 Plastic Surgery Milestones represented 52% (341 of 650) of questions, and 3 were not included in the ITE. Of 41 Orthopaedic Surgery Milestones, 7 represented 51% (201 of 394) of questions, and 5 had no representation on the ITE. Among the competencies, patient care was the most common (PSITE = 62% [403 of 650]; OITE = 59% [233 of 394]), followed by medical knowledge (PSITE = 34% [222 of 650]; OITE = 31% [124 of 394]). Distribution of the remaining competencies differed between the 2 specialties (PSITE = 4% [25 of 650]; OITE = 9% [37 of 394]).

CONCLUSIONS:
The ITEs tested slightly more than half of the milestones for the 2 specialties, and focused predominantly on patient care and medical knowledge competencies.
The Effect and Use of Milestones in the Assessment of Neurological Surgery Residents and Residency Programs.


OBJECTIVES:
The purpose of this study was to determine the effect of the Accreditation Council for Graduate Medical Education Milestones on the assessment of neurological surgery residents. The authors sought to determine the feasibility, acceptability, and utility of this new framework in making judgments of progressive competence, its implementation within programs, and the influence on curricula. Residents were also surveyed to elicit the effect of Milestones on their educational experience and professional development.

DESIGN, SETTING, AND PARTICIPANTS:
In 2015, program leadership and residents from 21 neurological surgery residency programs participated in an online survey and telephone interview in which they reflected on their experiences with the Milestones. Survey data were analyzed using descriptive statistics. Interview transcripts were analyzed using grounded theory.

RESULTS:
Response themes were categorized into 2 groups: outcomes of the Milestones implementation process, and facilitators and barriers. Because of Milestones implementation, participants reported changes to the quality of the assessment process, including the ability to identify struggling residents earlier and design individualized improvement plans. Some programs revised their curricula based on training gaps identified using the Milestones. Barriers to implementation included limitations to the adoption of a developmental progression model in the context of rotation block schedules and misalignment between progression targets and clinical experience. The shift from time-based to competency-based evaluation presented an ongoing adjustment for many programs. Organized preparation before clinical competency committee meetings and diverse clinical competency committee composition led to more productive meetings and perceived improvement in promotion decisions.

CONCLUSIONS:
The results of this study can be used by program leadership to help guide further implementation of the Milestones and program improvement. These results also help to guide the evolution of Milestones language and their implementation across specialties.
Multi-Institutional Studies: Quantitative

Education Research: The current state of neurophysiology education in selected neurology residency programs.


OBJECTIVE:
Prior research has illustrated there is a knowledge gap in neurology residents' neurophysiology education (EEG and EMG), and we sought to understand whether this is still an issue and to recognize the barriers in order to create solutions and improve education.

METHODS:
Surveys were developed for adult neurology residents and one for program directors asking about confidence in neurophysiology knowledge, percent of graduates reaching level 4 ACGME (American Council of Graduate Medical Education) milestones in EEG and EMG, methods of learning used, interest in the subjects, and suggestions for improvements.

RESULTS:
Twenty-six program directors (19% responder rate) and 55 residents (from at least 16 different programs) completed the survey. Program directors thought that 85% of graduating residents met level 4 milestones in EEG and only 75% in EMG. Structured rotations and more time allocated to education of these topics were frequent barriers mentioned. Postgraduate year 4 residents were 60% and 67% confident in EEG and 64%, 59%, and 62.3% in EMG level 4 milestones. Learning to read EEGs was considered important throughout residents' training; however, this interest and value decreased over time with EMG.

CONCLUSION:
In our study, program directors suspect up to a quarter of residents may graduate not meeting level 4 ACGME milestones, and residents expressed lack of confidence in these areas. The educational methods used to instruct residents in EEG and EMG were similar as were the barriers they face across programs. This information hopefully will help fuel curriculum design and interest in these important neurology techniques.


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.
Interim Analysis of a Prospective Multi-Institutional Study of Surgery Resident Experience with Flexibility in Surgical Training.


BACKGROUND:
The Flexibility in Surgical Training (FIST) consortium project was designed to evaluate the feasibility and resident outcomes of optional subspecialty-focused training within general surgery residency training.

STUDY DESIGN:
After approval by the American Board of Surgery, R4 and R5 residents were permitted to customize up to 12 of the final 24 months of residency for early tracking into 1 of 9 subspecialty tracks. A prospective IRB-approved study was designed across 7 institutions to evaluate the impact of this option on operative experience, in-service exam (American Board of Surgery In-Training Examination [ABSITE]) and ACGME milestone performance, and resident and program director (PD) perceptions. The FIST residents were compared with chief residents before FIST initiation (controls) as well as residents during the study period who did not participate in FIST (no specialization track, NonS).

RESULTS:
From 2013 to 2017, 122 of 214 chief residents (57%) completed a FIST subspecialty track. There were no differences in median ABSITE scores between FIST, NonS residents, and controls. The ACGME milestones at the end of the R5 year favored the FIST residents in 13 of 16 milestones compared with NonS. Case logs demonstrated an increase in track-specific cases compared with NonS residents. Resident and PD surveys reported a generally favorable experience with FIST.

CONCLUSIONS:
In this prospective study, FIST is a feasible option in participating institutions. All FIST residents, regardless of track, met requirements for ABS Board eligibility, despite modifications to rotations and case experience. Future studies will assess the impact of FIST on ABS exam results and fellowship success.
Identifying Gaps in the Performance of Pediatric Trainees Who Receive Marginal/Unsatisfactory Ratings.


PURPOSE:
To perform a derivation study to determine in which subcompetencies marginal/unsatisfactory pediatric residents had the greatest deficits compared with their satisfactorily performing peers and which subcompetencies best discriminated between marginal/unsatisfactory and satisfactorily performing residents.

METHOD:
Multi-institutional cohort study of all 21 milestones (rated on four or five levels) reported to the Accreditation Council for Graduate Medical Education, and global marginal/unsatisfactory versus satisfactory performance reported to the American Board of Pediatrics. Data were gathered in 2013-2014. For each level of training (postgraduate year [PGY] 1, 2, and 3), mean differences between milestone levels of residents with marginal/unsatisfactory and satisfactory performance adjusted for clustering by program and C-statistics (area under receiver operating characteristic curve) were calculated. A Bonferroni-corrected significance threshold of .0007963 was used to account for multiple comparisons.

RESULTS:
Milestone and overall performance evaluations for 1,704 pediatric residents in 41 programs were obtained. For PGY1s, two subcompetencies had almost a one-point difference in milestone levels between marginal/unsatisfactory and satisfactory trainees and outstanding discrimination (≥ 0.90): organize/prioritize (0.93; C-statistic: 0.91) and transfer of care (0.97; C-statistic: 0.90). The largest difference between marginal/unsatisfactory and satisfactory PGY2s was trustworthiness (0.78). The largest differences between marginal/unsatisfactory and satisfactory PGY3s were ethical behavior (1.17), incorporating feedback (1.03), and professionalization (0.96). For PGY2s and PGY3s, no subcompetencies had outstanding discrimination.

CONCLUSIONS:
Marginal/unsatisfactory pediatric residents had different subcompetency gaps at different training levels. While PGY1s may have global deficits, senior residents may have different performance deficiencies requiring individualized counseling and targeted performance improvement plans.
A Multicenter Prospective Comparison of the Accreditation Council for Graduate Medical Education Milestones: Clinical Competency Committee vs. Resident Self-Assessment.


OBJECTIVE:
The Accreditation Council for Graduate Medical Education requires accredited residency programs to implement competency-based assessments of medical trainees based upon nationally established Milestones. Clinical competency committees (CCC) are required to prepare biannual reports using the Milestones and ensure reporting to the Accreditation Council for Graduate Medical Education. Previous research demonstrated a strong correlation between CCC and resident scores on the Milestones at 1 institution. We sought to evaluate a national sampling of general surgery residency programs and hypothesized that CCC and resident assessments are similar.

DESIGN:
Details regarding the makeup and process of each CCC were obtained. Major disparities were defined as an absolute mean difference of ≥0.5 on the 4-point scale. A negative assessment disparity indicated that the residents evaluated themselves at a lower level than did the CCC. Statistical analysis included Wilcoxon rank sum and Sign tests.

SETTING:
CCCs and categorical general surgery residents from 15 residency programs completed the Milestones document independently during the spring of 2016.

RESULTS:
Overall, 334 residents were included; 44 (13%) and 43 (13%) residents scored themselves ≥0.5 points higher and lower than the CCC, respectively. Female residents scored themselves a mean of 0.08 points lower, and male residents scored themselves a mean of 0.03 points higher than the CCC. Median assessment differences for postgraduate year (PGY) 1-5 were 0.03 (range: -0.94 to 1.28), -0.11 (range: -1.22 to 1.22), -0.08 (range: -1.28 to 0.81), 0.02 (range: -0.91 to 1.00), and -0.19 (range: -1.16 to 0.50), respectively. Residents in university vs. independent programs had higher rates of negative assessment differences in medical knowledge (15% vs. 6%; P = 0.015), patient care (17% vs. 5%; P = 0.002), professionalism (23% vs. 14%; P = 0.013), and system-based practice (18% vs. 9%; P = 0.031) competencies. Major assessment disparities by sex or PGY were similar among individual competencies.

CONCLUSIONS:
Surgery residents in this national cohort demonstrated self-awareness when compared to assessments by their respective CCCs. This was independent of program type, sex, or level of training. PGY 5 residents, female residents, and those from university programs consistently rated themselves lower than the CCC, but these were not major disparities and the significance of this is unclear.
OBJECTIVE:
The Next Accreditation System implemented 5 levels of milestones for orthopedic surgery residents in 2013. The Level 1 milestones were noted as those "expected of an incoming resident." While the milestones were intended for assessing resident progression and readiness for independent practice, this designation can also be used to assess how well prepared graduating medical students are for beginning an orthopedic surgery residency. The primary objective of this paper is to measure recent medical school graduate comfort with the Level 1 milestones.

DESIGN, SETTING, AND PARTICIPANTS:
In June 2015, the program directors for the Midwest Orthopaedic Surgical Skills (MOSS) Consortium affiliated residency programs were sent an online survey for distribution to the recent medical school graduates who matched at their respective programs. The survey was about recent graduate comfort with the Level 1 milestone handles associated with 16 orthopedic milestones spanning multiple subspecialties. Responses were grouped based on comfort with individual milestone handles with orthopedic conditions (e.g., carpal tunnel) or with broader categories spanning orthopedic milestones (e.g., imaging).

RESULTS:
In all, 66 of 112 graduates (58.9%) responded. Of 60 milestone handles surveyed, respondents were "Comfortable" with an average of 31.6 ± 14.2 handles with some conditions performing much better than others. The median "Comfortable" response rate was 31 handles. The 8 broader categories had "Comfortable" response rates between 35% and 70%. All 8 orthopedic conditions had significantly higher "Comfortable" response rates for "Evaluation & Knowledge" handles than for "Decision Making & Treatment" handles.

CONCLUSIONS:
Most recent medical student graduates who matched into an orthopedic surgery residencies are only comfortable with about half of the Level 1 milestone handles even though they are expected to meet the Level 1 milestones upon beginning residency. This finding suggests the development of an assessment based on the Level 1 milestones would be appropriate to better inform both graduate and undergraduate medical education in orthopedic surgery.
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.
Validity Evidence From Ratings of Pediatric Interns and Subinterns on a Subset of Pediatric Milestones.


PURPOSE:
To investigate evidence for validity of faculty members' pediatric milestone (PM) ratings of interns (first-year residents) and subinterns (fourth-year medical students) on nine subcompetencies related to readiness to serve as a pediatric intern in the inpatient setting.

METHOD:
The Association of Pediatric Program Directors Longitudinal Educational Assessment Research Network (APPD LEARN) and the National Board of Medical Examiners collaborated to investigate the utility of assessments of the PMs for trainees' performance. Data from 32 subinterns and 179 interns at 17 programs were collected from July 2012 through April 2013. Observers used several tools to assess learners. At each site, a faculty member used these data to make judgments about the learner's current developmental milestone in each subcompetency. Linear mixed models were fitted to milestone judgments to examine their relationship with learner's rank and subcompetency.

RESULTS:
On a 5-point developmental scale, mean milestone levels for interns ranged from 3.20 (for the subcompetency Work effectively as a member of a team) to 3.72 (Humanism) and for subinterns from 2.89 (Organize and prioritize care) to 3.61 (Professionalization). Mean milestone ratings were significantly higher for the Professionalism competency (3.59-3.72) for all trainees compared with Patient Care (2.89-3.24) and Personal and Professional Development (3.33-3.51). Mean intern ratings were significantly higher than mean subintern ratings for all nine subcompetencies except Professionalization, Humanism, and Trustworthiness.

CONCLUSIONS:
The PMs had a coherent internal structure and could distinguish between differing levels of trainees, which supports their validation for documenting developmental progression of pediatric trainees.
Comparison of Male vs Female Resident Milestone Evaluations by Faculty During Emergency Medicine Residency Training.


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 PARTICIPANTS:
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.
Competent for Unsupervised Practice: Use of Pediatric Residency Training Milestones to Assess Readiness.


PURPOSE:
To describe clinical skills progression during pediatric residency using the distribution of pediatric milestone assessments by subcompetency and year of training and to determine reasonable milestone expectations at time of graduation.

METHOD:
Multi-institutional cohort study of the milestones reported to the Accreditation Council for Graduate Medical Education for all 21 pediatric subcompetencies. Most subcompetencies were measured using five milestone levels (1 = novice, 2 = advanced beginner, 3 = competent, 4 = proficient, 5 = master); 3 subcompetencies had only four levels defined.

RESULTS:
Milestone assessments for 2,030 pediatric residents in 47 programs during academic year 2013-2014 were obtained. There was significant variation in end-of-year milestone ratings for residents within each level of training, which decreased as training level increased. Most (78.9%; 434/550) graduating third-year pediatric residents received a milestone rating of ≥ 3 in all 21 subcompetencies; fewer (21.1%; 116/550) received a rating of ≥ 4 in all subcompetencies. Across all training levels, professionalism and interpersonal communication skills were rated highest; quality improvement was rated lowest.

CONCLUSIONS:
Trainees entered residency with a wide range of skills. As they advanced, skill variability within a training level decreased. Most graduating pediatric residents were still advancing on the milestone continuum toward proficiency and mastery, and an expectation of milestone ratings ≥ 4 in all categories upon graduation is unrealistic; milestone ratings ≥ 3 upon graduation may be more realistic. Understanding current pediatric residents’ and graduates’ skills can help to identify key areas that should be specifically targeted during training.
A taxonomy of perioperative surgical learning: Trending resident skill acquisition.


BACKGROUND:
Resident and curriculum evaluation require tracking surgical resident operative performance, yet what and when to measure remains unclear.

METHODS:
From a multi-institutional database, we reviewed 611 resident/surgeon-paired assessments of ACGME Milestones and modified OPRS ratings for different cases and postgraduate years.

RESULTS:
Faculty Milestone ratings increased with each PGY (p=<0.001) and correlated with resident self-ratings (ICC = 0.83). Mean OPRS scores increased in small increments with substantial intra-year variability. Progression among individual OPRS subcategories was not apparent from more global analyses. Interestingly, male faculty offered lower ratings than female faculty.

CONCLUSIONS:
Milestones and modified mean OPRS ratings suggest residents are learning, yet lack sufficient discrimination for promotion or curricular analysis. Differential progression through OPRS subcategories suggests a taxonomy of surgical learning that can be tailored to focus on different skills at each point in the training continuum. The effect of faculty gender on resident ratings awaits further study.
INTRODUCTION:
Minimally invasive surgery has become an important aspect of Pediatric Urology fellowship training. In 2014, the Accreditation Council for Graduate Medical Education published the Pediatric Urology Milestone Project as a metric of fellow proficiency in multiple facets of training, including laparoscopic/robotic procedures.

OBJECTIVE:
The present study assessed trends in minimally invasive surgery training and utilization of the Milestones among recent Pediatric Urology fellows.

STUDY DESIGN:
Using an electronic survey instrument, Pediatric Urology fellowship program directors and fellows who completed their clinical year in 2015 were surveyed. Participants were queried regarding familiarity with the Milestone Project, utilization of the Milestones, robotic/laparoscopic case volume and training experience, and perceived competency with robotic/laparoscopic surgery at the start and end of the fellowship clinical year according to Milestone criteria. Responses were accepted between August and November 2015.

RESULTS:
Surveys were distributed via e-mail to 35 fellows and 30 program directors. Sixteen fellows (46%) and 14 (47%) program directors responded. All fellows reported some robotic experience prior to fellowship, and 69% performed >50 robotic/laparoscopic surgeries during residency. Fellow robotic/laparoscopic case volume varied: three had 1-10 cases (19%), four had 11-20 cases (25%), and nine had >20 cases (56%). Supplementary or robotic training modalities included simulation (9), animal models (6), surgical videos (7), and courses (2). Comparison of beginning and end of fellowship robotic/laparoscopic Milestone assessment (Summary Fig.) revealed scores of <3 in (10) 62% of fellow self-assessments and 10 (75%) of program director assessments. End of training Milestone scores >4 were seen in 12 (75%) of fellow self-assessment and eight (57%) of program director assessments.

DISCUSSION:
An improvement in robotic/laparoscopic Milestone scores by both fellow self-assessment and program director assessment was observed during the course of training; however, 43% of program directors rated their fellow below the graduation target of a Milestone score of 4.

CONCLUSION:
The best ways to teach minimally invasive surgery in fellowship training must be critically considered.
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.
Have First-Year Emergency Medicine Residents Achieved Level 1 on Care-Based Milestones?

BACKGROUND:
Residents in Accreditation Council for Graduate Medical Education accredited emergency medicine (EM) residencies were assessed on 23 educational milestones to capture their progression from medical student level (Level 1) to that of an EM attending physician (Level 5). Level 1 was conceptualized to be at the level of an incoming postgraduate year (PGY)-1 resident; however, this has not been confirmed.

OBJECTIVES:
Our primary objective in this study was to assess incoming PGY-1 residents to determine what percentage achieved Level 1 for the 8 emergency department (ED) patient care-based milestones (PC 1-8), as assessed by faculty. Secondary objectives involved assessing what percentage of residents had achieved Level 1 as assessed by themselves, and finally, we calculated the absolute differences between self- and faculty assessments.

METHODS:
Incoming PGY-1 residents at 4 EM residencies were assessed by faculty and themselves during their first month of residency. Performance anchors were adapted from ACGME milestones.

RESULTS:
Forty-one residents from 4 programs were included. The percentage of residents who achieved Level 1 for each subcompetency on faculty assessment ranged from 20% to 73%, and on self-assessment from 34% to 92%. The majority did not achieve Level 1 on faculty assessment of milestones PC-2, PC-3, PC-5a, and PC-6, and on self-assessment of PC-3 and PC-5a. Self-assessment was higher than faculty assessment for PC-2, PC-5b, and PC-6.

CONCLUSIONS:
Less than 75% of PGY-1 residents achieved Level 1 for ED care-based milestones. The majority did not achieve Level 1 on 4 milestones. Self-assessments were higher than faculty assessments for several milestones.
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.
Internal Medicine Residents’ Perspectives on Receiving Feedback in Milestone Format.

BACKGROUND:
In contrast to historical feedback, which was vague or provided residents’ numerical scores without clear meaning, milestone-based feedback is focused on specific knowledge, skills, and behaviors that define developmental trajectory. It was anticipated that residents would welcome the more specific and actionable feedback provided by the milestone framework, but this has not been studied.

OBJECTIVE:
We assessed internal medicine (IM) residents’ perceptions of receiving feedback in the milestone framework, particularly assessing perception of the utility of milestone-based feedback compared to non-milestone-based feedback.

METHODS:
We surveyed a total of 510 IM residents from 7 institutions. Survey questions assessed resident perception of milestone feedback in identifying strengths, weaknesses, and trajectory of professional development. Postgraduate years 2 and 3 (PGY-2 and PGY-3) residents were asked to compare milestones with prior methods of feedback.

RESULTS:
Of 510 residents, 356 (69.8%) responded. Slightly less than half of the residents found milestone-based feedback “extremely useful” or “very useful” in identifying strengths (44%), weaknesses (43%), specific areas for improvement (45%), and appropriate education progress (48%). Few residents found such feedback "not very useful" or "not at all useful" in these domains. A total of 51% of PGY-2 and PGY-3 residents agreed that receiving milestone-based feedback was more helpful than previous forms of feedback.

CONCLUSIONS:
IM residents are aware of the concepts of milestones, and half of the residents surveyed found milestone feedback more helpful than previous forms of feedback. More work needs to be done to understand how milestone-based feedback could be delivered more effectively to enhance resident development.
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 α = 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.
Multi-Institutional Studies: Qualitative

Milestones for the Final Mile: Interspecialty Distinctions in Primary Palliative Care Skills Training.


CONTEXT:
Primary palliative care (PPC) skills are useful in a wide variety of medical and surgical specialties, and the expectations of PPC skill training are unknown across graduate medical education.

OBJECTIVES:
We characterized the variation and quality of PPC skills in residency outcomes-based Accreditation Council for Graduate Medical Education (ACGME) milestones.

METHODS:
We performed a content analysis with structured implicit review of 2015 ACGME milestone documents from 14 medical and surgical specialties chosen for their exposure to clinical situations requiring PPC. For each specialty milestone document, we characterized the variation and quality of PPC skills in residency outcomes-based ACGME milestones.

RESULTS:
We identified 959 occurrences of 29 palliative search terms within 14 specialty milestone documents. Within these milestone documents, implicit review characterized 104 milestones with direct saliency to PPC skills and 196 milestones with indirect saliency. Initial interrater agreement of the saliency rating among the primary reviewers was 89%. Specialty milestone documents varied widely in their incorporation of PPC skills within milestone documents. PPC milestones were most commonly found in milestone documents for Anesthesiology, Pediatrics, Urology, and Physical Medicine and Rehabilitation. PPC-relevant milestones were most commonly found in the Interpersonal and Communication Skills core competency with 108 (36%) relevant milestones classified under this core competency.

CONCLUSIONS:
Future revisions of specialty-specific ACGME milestone documents should focus on currently underrepresented, but important PPC skills.
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.
Reviewing residents' competence: a qualitative study of the role of clinical competency committees in performance assessment.


PURPOSE:
Clinical competency committees (CCCs) are now required in graduate medical education. This study examined how residency programs understand and operationalize this mandate for resident performance review.

METHOD:
In 2013, the investigators conducted semistructured interviews with 34 residency program directors at five public institutions in California, asking about each institution's CCCs and resident performance review processes. They used conventional content analysis to identify major themes from the verbatim interview transcripts.

RESULTS:
The purpose of resident performance review at all institutions was oriented toward one of two paradigms: a problem identification model, which predominated; or a developmental model. The problem identification model, which focused on identifying and addressing performance concerns, used performance data such as red-flag alerts and informal information shared with program directors to identify struggling residents. In the developmental model, the timely acquisition and synthesis of data to inform each resident's developmental trajectory was challenging. Participants highly valued CCC members' expertise as educators to corroborate the identification of struggling residents and to enhance credibility of the committee's outcomes. Training in applying the milestones to the CCC's work was minimal. Participants were highly committed to performance review and perceived the current process as adequate for struggling residents but potentially not for others.

CONCLUSIONS:
Institutions orient resident performance review toward problem identification; a developmental approach is uncommon. Clarifying the purpose of resident performance review and employing efficient information systems that synthesize performance data and engage residents and faculty in purposeful feedback discussions could enable the meaningful implementation of milestones-based assessment.
Early feedback on the use of the internal medicine reporting milestones in assessment of resident performance.


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.
The pediatric milestones: initial evidence for their use as learning road maps for residents.


OBJECTIVE:
As the next step in competency-based medical education, the Pediatrics Milestone Project seeks to provide a learner-centered approach to training and assessment. To help accomplish this goal, this study sought to determine how pediatric residents understand, interpret, and respond to the Pediatrics Milestones.

METHODS:
Cognitive interviews with 48 pediatric residents from all training levels at 2 training programs were conducted. Each participant reviewed one Pediatrics Milestone document (PMD). Eight total Pediatrics Milestones, chosen for their range of complexity, length, competency domain, and primary author, were included in this study. Six residents, 2 from each year of residency training, reviewed each PMD. Interviews were transcribed and coded using inductive methods, and codes were grouped into themes that emerged.

RESULTS:
Four major themes emerged through coding and analysis: 1) the participants' degree of understanding of the PMDs is sufficient, often deep; 2) the etiology of participants' understanding is rooted in their experiences; 3) there are qualities of the PMD that may contribute to or detract from understanding; and 4) participants apply their understanding by noting the PMD describes a developmental progression that can provide a road map for learning. Additionally, we learned that residents are generally comfortable being placed in the middle of a series of developmental milestones. Two minor themes focusing on interest and practicality were also identified.

CONCLUSIONS:
This study provides initial evidence for the Pediatrics Milestones as learner-centered documents that can be used for orientation, education, formative feedback, and, ultimately, assessment.
Single Institution Studies: Quantitative

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.
Long-term Retention of Musculoskeletal Ultrasound Training During Residency.


Abstract
The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Physical Medicine and Rehabilitation (ABPMR) developed milestones for evaluation of resident physicians that include proper musculoskeletal ultrasound (MSUS) examination of major joints. To date, there have been no published data demonstrating acquisition and retention of these skills and correlation with the milestone evaluation. The investigators developed and implemented a curriculum in musculoskeletal ultrasound examination for Physical Medicine and Rehabilitation (PM&R) residents at a large academic medical center. The investigators chose six joints for training and evaluation: ankle, elbow, hip, knee, shoulder and wrist/hand. The program included: 1) didactic lectures on anatomy and ultrasound technique; 2) peer-led demonstrations of the procedure on a standardized patient (SP); 3) individual practice on SPs; 4) faculty observation and feedback; 5) review sessions and additional practice; and, 6) assessment of skills in an objective structured clinical examination (OSCE). From 2013-2017, 30 PM&R residents were trained and evaluated. The results, based on OSCE scores, showed that the majority of residents achieved the appropriate level of competency for their year. A blended, standardized curriculum in MSUS instruction with assessment by an OSCE, can be used to evaluate MSUS skills, and can help align this education with residency milestones.
Using an alumni survey to target improvements in an emergency medicine training programme.


INTRODUCTION:
The Accreditation Council for Graduate Medical Education (ACGME) is the governing body responsible for accrediting graduate medical training programme in the USA. The Emergency Medicine Milestones (EM-Milestones) were developed by the ACGME and American Board of Emergency Medicine as a guide and monitoring tool for the knowledge, skills, abilities and experiences to be acquired during training. Alumni surveys have been reported as a valuable resource for training programme to identify areas for improvement; however, there are few studies regarding programme improvement in emergency medicine. We aimed to use the EM-Milestones, adapted as an alumni self-assessment survey, to identify areas for training programme improvement.

METHODS:
This study was conducted at an urban, academic affiliated, community hospital in New York city with an emergency medicine training programme consisting of 30 residents of our emergency medicine training programme were sent an EM-Milestones-based self-assessment survey. Participants evaluated their ability in each EM-Milestones subcompetency on a Likert scale. Data were analysed using descriptive statistics.

RESULTS:
Response rate was 74% (69/93). Alumni reported achieving the target performance in 5/6 general competencies, with Systems-Based Practice falling below the target performance. The survey further identified 6/23 subcompetencies (Pharmacotherapy, Ultrasound, Wound Management, Patient Safety, Systems-Based Management and Technology) falling below the target performance level.

DISCUSSION:
Alumni self-evaluation of competence using the EM-Milestones provides valuable information concerning confidence to practice independently; these data, coupled with regular milestone evaluation of existing trainees, can identify problem areas and provide a blueprint for targeted programme improvement.
Burnout is Associated With Emotional Intelligence but not Traditional Job Performance Measurements in Surgical Residents.


OBJECTIVE: To evaluate whether burnout was associated with emotional intelligence and job performance in surgical residents.

DESIGN: General surgery residents at a single institution were surveyed using the Maslach Burnout Inventory (MBI) and trait EI questionnaire (TEIQ-SF). Burnout was defined as scoring in 2 of the 3 following domains: Emotional Exhaustion (high), Depersonalization (high), and Personal Accomplishment (low). Job performance was evaluated using faculty evaluations of clinical competency-based surgical milestones and standardized test scores including the American Board of Surgery In-Training Exam (ABSITE) and the United States Medical Licensing Examination (USMLE) Step 3. USMLE Step 1 and USMLE Step 2, which were taken prior to residency training, were included to examine possible associations of burnout with USMLE examinations. Statistical comparison was made using Pearson correlation and simple linear regression adjusting for PGY level.

SETTING: This study was conducted at the University of Alabama at Birmingham (UAB) general surgery residency program.

PARTICIPANTS: All current and incoming general surgery residents at UAB were invited to participate in this study.

RESULTS: Forty residents participated in the survey (response rate 77%). Ten residents, evenly distributed from incoming residents to PGY-4, had burnout (25%). Mean global EI was lower in residents with burnout versus those without burnout (3.71 vs 3.9, p = 0.02). Of the 4 facets of EI, mean self-control values were lower in residents with burnout versus those without burnout (3.3 vs 4.06, p < 0.01). Each component of burnout was associated with global EI, with the strongest correlation being with personal accomplishment (r = 0.64; p < 0.01). Residents with burnout did not have significantly different mean scores for USMLE Step 1 (229 vs 237, p = 0.12), Step 2 (248 vs 251, p = 0.56), Step 3 (223 vs 222, p = 0.97), or ABSITE percentile (44.6 vs 58, p = 0.33) compared to residents without burnout. Personal accomplishment was associated with ABSITE percentile scores (r = 0.35; p = 0.049). None of the 16 surgical milestone scores were significantly associated with burnout.

CONCLUSIONS: Burnout is present in surgery residents and associated with emotional intelligence. There was no association of burnout with USMLE scores, ABSITE percentile, or surgical milestones. Traditional methods of assessing resident performance may not be capturing burnout and strategies to reduce burnout should consider targeting emotional intelligence.
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).
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 = 0.70, 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.
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.
Are Surgical Milestone Assessments Predictive of In-Training Examination Scores?


OBJECTIVES:
With the recent utilization of Accreditation Council for Graduate Medical Education developmental milestones as part of resident evaluation, we sought to explore whether milestone-based ratings were associated with American Board of Surgery In-Training Examination (ABSITE) scores.

METHODS:
Mid-year milestone ratings were obtained from the Accreditation Council for Graduate Medical Education Accreditation Database System for years 2014, 2015, and 2016 for all postgraduate years 1-5 general surgery residents in our program and paired with ABSITE scores (n = 69) from January of the following year. Linear regression was used to assess predictor importance of milestones on both ABSITE percentage correct scores and ABSITE percentile scores.

RESULTS:
Minimal to small positive correlations were observed between milestones and ABSITE percentile scores (r = 0.09-0.25), while moderately large correlations were observed between milestones and percentage correct scores (r = 0.65-0.76). The Medical Knowledge 1 (MK1) milestone significantly predicted ABSITE percentage correct scores, and explained 60% of the variance (adjusted R² = 0.603). MK1 also significantly predicted ABSITE percentile scores, although weaker in magnitude, with MK1 explaining 20% of the variance (adjusted R² = 0.197). Postgraduate year level and other milestones were not influential predictors of ABSITE scores.

CONCLUSIONS:
The mid-year MK1 milestone rating was predictive of ABSITE scores and may serve as a useful marker for Clinical Competency Committees to identify residents who could benefit from additional support to prepare for the ABSITE, although given the small exploratory nature of this study, additional research is still needed.
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).


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.
A competency-based simulation curriculum for surgical resident trauma resuscitation skills.


BACKGROUND:
Evidence-based curricula for nonprocedural simulation training in general surgery are lacking. Residency programs are required to implement simulation training despite this shortcoming. The goal of this project was the development of a simulation curriculum that measurably improves milestone performance and replaces traditional experienced-based training with a competency-based model.

MATERIALS AND METHODS:
SimMan 3G® (Laerdal Medical, Wappingers Falls, NY, USA) was utilized for simulation. Needs assessment targeted trauma and shock resuscitation. Scenario design applied deliberate practice methodology. Learner performance data included items such as identification of shock physiology, resuscitation products used, volume delivered, use of resuscitation end-points, and knowledge of massive transfusion. Characteristics essential for a successful program were tabulated.

RESULTS:
Forty-eight residents in postgraduate year (PGY) 2-5 participated representing 100% of the 48 eligible for the training. Senior residents (PGY 4 and 5) demonstrated near universal improvement. Junior residents (PGY 2 and 3) improved in some areas but showed more skill decay between sessions. Overall, milestone performance improved with each training session, and resident feedback was universally positive.

CONCLUSIONS:
This prototype curriculum improved surgical resident competency in shock resuscitation in a simulated patient care environment. It can be modified to accommodate centers with fewer resources and can be implemented by clinical faculty. The essential characteristics of a successful program are identified.
Towards Consensus: Training in Procedural Skills for Diagnostic Radiology Residents- Current Opinions of Residents and Faculty at a Large Academic Center.


RATIONALE AND OBJECTIVES:
The Diagnostic Radiology Milestones Project provides a framework for measuring resident competence in radiologic procedures, but there are limited data available to assist in developing these guidelines. We performed a survey of current radiology residents and faculty at our institution as a first step toward obtaining data for this purpose. The survey addressed attitudes toward procedural standardization and procedures that trainees should be competent by the end of residency.

MATERIALS AND METHODS:
Current residents and faculty members were surveyed about whether or not there should be standardization of procedural training, in which procedures residents should achieve competency, and the number of times a procedure needs to be performed to achieve competency.

RESULTS:
Survey data were received from 60 study participants with an overall response rate of 32%. Sixty-five percent of respondents thought that procedural training should be standardized. Standardization of procedural training would include both the list of procedures that trainees should be competent in at the end of residency and the standard minimum number of procedures to achieve competency. Procedures that both residents and faculty agreed are important in which to achieve competency included central line/port procedures; CT-guided abdominal, thoracic, and musculoskeletal procedures; minor fluoroscopic-guided procedures; general fluoroscopy; peripheral line placements; and US-guided abdominal procedures. For most of these categories, most respondents believed that these procedures needed to be performed 6-20 times to achieve competency.

CONCLUSION:
Both resident and faculty respondents agreed that procedural training should be standardized during residency, and competence in specific procedures should be achieved at the completion of residency. Although this study is limited to a single institution, our data may provide assistance in developing future guidelines for standardizing image-guided procedure training. Future studies could be expanded to create a national consensus regarding the implementation of the Diagnostic Radiology Milestones Project.
Critical Deficiency Ratings in Milestone Assessment: A Review and Case Study.


PURPOSE:
The Accreditation Council for Graduate Medical Education (ACGME) requires programs to report learner progress using specialty-specific milestones. It is unclear how milestones can best identify critical deficiencies (CDs) in trainee performance. Specialties developed milestones independently of one another; not every specialty included CDs within milestones ratings. This study examined the proportion of ACGME milestone sets that include CD ratings, and describes one residency program's experiences using CD ratings in assessment.

METHOD:
The authors reviewed ACGME milestones for all 99 specialties in November 2015, determining which rating scales contained CDs. The authors also reviewed three years of data (July 2012-June 2015) from the University of Cincinnati Medical Center (UCMC) internal medicine residency assessment system based on observable practice activities mapped to ACGME milestones. Data were analyzed by postgraduate year, assessor type, rotation, academic year, and core competency. The Mantel-Haenszel chi-square test was used to test for changes over time.

RESULTS:
Specialties demonstrated heterogeneity in accounting for CDs in ACGME milestones, with 22% (22/99) of specialties having no language describing CDs in milestones assessment. Thirty-three percent (63/189) of UCMC internal medicine residents received at least one CD rating, with CDs accounting for 0.18% (668/364,728) of all assessment ratings. The authors identified CDs across multiple core competencies and rotations.

CONCLUSIONS:
Despite some specialties not accounting for CDs in milestone assessment, UCMC's experience demonstrates that a significant proportion of residents may be rated as having a CD during training. Identification of CDs may allow programs to develop remediation and improvement plans.
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.


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 ($\text{weighted Cohen's } \kappa = 0.72$, $p < 0.001$). 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.
Initial Comparison of Resident and Attending Milestones Evaluations in Plastic Surgery.


BACKGROUND:
Graduate medical education has recently undergone a major archetypal shift toward competency-based evaluations of residents' performance. The implementation of the Milestones program by the Accreditation Council for Graduate Medical Education (ACGME) is a core component of the shift, designed to ensure uniformity in measuring residency knowledge using a series of specialty-specific achievements. This study evaluates the correlation between residents' self-evaluations and program directors' assessments of their performance.

METHODS:
The study population comprised 12 plastic surgery residents, ranging from postgraduate year 1 to postgraduate year 6, enrolled in an integrated residency program at a single institution.

RESULTS:
Overall, average attending scores were lower than average resident scores at all levels except postgraduate year 6. Correlation between resident and attending evaluations ranged from 0.417 to 0.957, with the correlation of average scores of Patient Care (0.854) and Medical Knowledge (0.816) Milestones significantly higher than those of professional skillsets (0.581). "Patient care, facial esthetics" was the Milestone with the lowest average scores from both groups. Residents scored themselves notably higher than their attendings' evaluations in Practice-based Learning and Improvement categories (+0.958) and notably lower in Medical Knowledge categories such as "Cosmetic Surgery, Trunk and Lower Extremities" (-0.375) and "Non-trauma hand" (-0.208). The total possible number of participants in this study was 12. The actual number of participants was 12 (100%).

CONCLUSIONS:
The remarkable range of correlations suggests that expectations for performance standards may vary widely between residents and program directors. Understanding gaps between expectations and performance is vital to inform current and future residents as the restructuring of the accreditation process continues.
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.
Milestones: a rapid assessment method for the Clinical Competency Committee.

INTRODUCTION:
Educational milestones are now used to assess the developmental progress of all U.S. graduate medical residents during training. Twice annually, each program’s Clinical Competency Committee (CCC) makes these determinations and reports its findings to the Accreditation Council for Graduate Medical Education (ACGME). The ideal way to conduct the CCC is not known. After finding that deliberations reliant upon the new milestones were time intensive, our internal medicine residency program tested an approach designed to produce rapid but accurate assessments.

MATERIAL AND METHODS:
For this study, we modified our usual CCC process to include pre-meeting faculty ratings of resident milestones progress with in-meeting reconciliation of their ratings. Data were considered largely via standard report and presented in a pre-arranged pattern. Participants were surveyed regarding their perceptions of data management strategies and use of milestones. Reliability of competence assessments was estimated by comparing pre-/post-intervention class rank lists produced by individual committee members with a master class rank list produced by the collective CCC after full deliberation.

RESULTS:
Use of the study CCC approach reduced committee deliberation time from 25 min to 9 min per resident ($p<0.001$). Committee members believed milestones improved their ability to identify and assess expected elements of competency development ($p = 0.026$). Individual committee member assessments of trainee progress agreed well with collective CCC assessments.

CONCLUSIONS:
Modification of the clinical competency process to include pre-meeting competence ratings with in-meeting reconciliation of these ratings led to shorter deliberation times, improved evaluator satisfaction and resulted in reliable milestone assessments.
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.
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 α 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 α: 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 their residency.
Do Attending Surgeons and Residents See Eye To Eye? An Evaluation of the Accreditation Council For Graduate Medical Education Milestones in General Surgery Residency.


OBJECTIVE:
The Accreditation Council for Graduate Medical Education requires accredited general surgery residencies to implement competency-based developmental outcomes in resident evaluations. Overall, 16 milestones are evaluated by a clinical competency committee (CCC). The milestones span 8 domains of surgical practice and 6 Accreditation Council for Graduate Medical Education clinical competencies.

The highest level suggests preparedness for independent practice. Our objective was to compare self-assessments and committee evaluations within the milestone framework.

STUDY DESIGN:
All residents underwent semiannual evaluations from 2013 to 2015. Residents independently completed a self-assessment using the milestones. The CCC completed the milestones document using resident evaluations and consensus opinion of committee members. Assessment differences were calculated for each evaluation. A negative value indicated that the residents evaluated themselves at a lower level than the committee. Major assessment disparities were defined as >0.5 on a 4-point scale.

SETTING:
An independent academic medical center.

PARTICIPANTS:
General surgery residents.

RESULTS:
Overall, 20 residents participated; 7 were female. In total, 5 (7%) evaluations had a mean overall assessment difference >0.5, whereas 6 (8%) had a difference <-0.5. Residents evaluated themselves lower than the committee with a median assessment difference of -0.06 [-0.25 to 0.16] (p = 0.041).

Evaluations were similar across surgical domains. Negative self-evaluations were more common for medical knowledge (-0.25 [-0.25 to 0.25], p = 0.025). Female residents had 2% positive and 13% negative major assessment disparity rates versus 10% positive and 9% negative rates among male residents.

Postgraduate year III residents had 12% positive and 4% negative major disparity rates; all other years had higher negative than positive rates.

CONCLUSIONS:
Surgery residents within our program demonstrated adequate self-awareness, with most self-evaluations falling within a half level of the CCC report. This self-awareness was consistent across surgical domains and most clinical competencies. Residents perceived a lower level of medical knowledge than the CCC. Subgroup analysis revealed interesting trends in the effects of sex, postgraduate year level, and academic year timing, which will take additional study to fully delineate.
Entrusting Observable Practice Activities and Milestones Over the 36 Months of an Internal Medicine Residency.


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.
A first look at the Accreditation Council for Graduate Medical Education anesthesiology milestones: implementation of self-evaluation in a large residency program.


STUDY OBJECTIVE:
The objective was to determine if there is a correlation between resident postgraduate year (PGY) of training and self-evaluation of performance using the Accreditation Council for Graduate Medical Education milestones.

DESIGN:
Survey.

SETTING:
Residency program at a large academic center.

PATIENTS:
Residents and Faculty Clinical Competency Committee (CCC).

INTERVENTIONS:
None.

MEASUREMENTS:
Resident and CCC milestone scores.

MAIN RESULTS:
Correlation coefficients for average score for each milestone vs PGY level ranged from 0.80 for receiving and giving feedback to 0.95 for anesthetic choice and conduct. All milestones showed a relatively linear relationship with PGY of training, and none were found to be consistently reached very late or very early in training. When examining variation across the scores for the individual residents, the distributions for PGY-2 and -3 appeared to be wider than those for PGY-1 and -4. The intraclass correlation coefficients ranged from 0.718 to 0.928.

CONCLUSIONS:
There was a remarkable degree of consistency in the relationship between level of training and resident self-assessment score for every milestone, as well as strong agreement between the resident and CCC faculty scores. Examination of the variance in the scores, when interpreted in light of our particular training program's characteristics, suggests that the milestones accurately reflect the progression in skill across the residency. In addition, given the concordance between the self-evaluation scores and the CCC faculty scores, self-evaluation may be a reasonable starting point as programs begin the daunting task of determining scores for each of the 25 milestones as part of the biannual evaluation process.
A Milestone-Based Evaluation System-The Cure for Grade Inflation?

PURPOSE:
Controversy exists over the optimal use of the Milestones in the process of resident evaluation and feedback. We sought to evaluate the performance of a Milestones-based feedback system in comparison to a traditional model.

METHODS:
The traditional evaluation system (TES) consisted of a generic 16-item survey using a 5-point Likert scale ranging from 1 to 5, and a free-text comments section. The Milestones-based evaluation system (MBES) was launched in July 2014, ranging from 0 to 4. Individual milestones were mapped to rotations based on resident educational goals by postgraduate year (PGY). The MBES consisted of a survey with a maximum of 7 items, followed by a free-text comment section. Within each evaluation system, an overall composite score was calculated for each categorical general surgical resident. To scale the 2 systems for comparison, TES scores were adjusted downward by 1 point. Descriptive statistics were performed. Univariate analysis was performed with the Wilcoxon signed-rank test. A test for trend across PGY was used for the MBES only.

RESULTS:
In the traditional system, the median score was 3.66 (range: 3.2-4.0). There was no meaningful difference in the median score by PGY. In the new system, the median score was 2.69 (range: 1.5-3.7, p < 0.01). The median score differed across PGY and increased by PGY of training (p < 0.01). There was an increase in differences between median scores by PGY.

CONCLUSIONS:
On using the milestones to facilitate faculty evaluation of resident knowledge and skill, there was a trend in increasing score by PGY of training. In the MBES, scores could be used to better discriminate resident skill and knowledge levels and resulted in improved differentiation in scoring by PGY. The use of the milestones as a basis for evaluation enabled the program to provide more meaningful feedback to residents and represents an improvement in surgical education.
Emergency Medicine Residents Consistently Rate Themselves Higher than Attending Assessments on ACGME Milestones.


INTRODUCTION:
In 2012 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which implemented milestones to assess the competency of residents and fellows. While attending evaluation and feedback is crucial for resident development, perhaps equally important is a resident's self-assessment. If a resident does not accurately self-assess, clinical and professional progress may be compromised. The objective of our study was to compare emergency medicine (EM) resident milestone evaluation by EM faculty with the same resident's self-assessment.

METHODS:
This is an observational, cross-sectional study that was performed at an academic, four-year EM residency program. Twenty-five randomly chosen residents completed milestone self-assessment using eight ACGME sub-competencies deemed by residency leadership as representative of core EM principles. These residents were also evaluated by 20 faculty members. The milestone levels were evaluated on a nine-point scale. We calculated the average difference between resident self-ratings and faculty ratings, and used sample t-tests to determine statistical significance of the difference in scores.

RESULTS:
Eighteen residents evaluated themselves. Each resident was assessed by an average of 16 attendings (min=10, max=20). Residents gave themselves statistically significant higher milestone ratings than attendings did for each sub-competency examined (p<0.0001).

CONCLUSION:
Residents over-estimated their abilities in every sub-competency assessed. This underscores the importance of feedback and assessment transparency. More attention needs to be paid to methods by which residency leadership can make residents' self-perception of their clinical ability more congruent with that of their teachers and evaluators. The major limitation of our study is small sample size of both residents and attendings.
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.
Milestone-Based Assessments Are Superior to Likert-Type Assessments in Illustrating Trainee Progression.


BACKGROUND: The Pediatrics Milestone Project uses behavioral anchors, narrative descriptions of observable behaviors, to describe learner progression through the Accreditation Council for Graduate Medical Education competencies. Starting June 2014, pediatrics programs were required to submit milestone reports for their trainees semiannually. Likert-type scale assessment tools were not designed to inform milestone reporting, creating a challenge for Clinical Competency Committees.

OBJECTIVE: To determine if milestone-based assessments better stratify trainees by training level compared to Likert-type assessments.

METHODS: We compared assessment results for 3 subcompetencies after changing from a 5-point Likert scale to milestone-based behavioral anchors in July 2013. Program leadership evaluated the new system by (1) comparing PGY-1 mean scores on Likert-type versus milestone-based assessments; and (2) comparing mean scores on the Likert-type versus milestone-based assessments across PGY levels.

RESULTS: Mean scores for PGY-1 residents were significantly higher on the prior year’s Likert-type assessments than milestone-based assessments for all 3 subcompetencies (P, .01). Stratification by PGY level was not observed with Likert-type assessments (eg, interpersonal and communication skills 1 [ICS1] mean score for PGY-1, 3.99 versus PGY-3, 3.98; P 5 .98). In contrast, milestone-based assessments demonstrated stratification by PGY level (eg, the ICS1 mean score was 3.06 for PGY-1, 3.83 for PGY-2, and 3.99 for PGY-3; P,.01 for PGY-1 versus PGY-3). Significantly different means by trainee level were noted across 21 subcompetencies on milestone-based assessments (P, .01 for PGY-1 versus PGY-3).

CONCLUSIONS: Initial results indicate milestone-based assessments stratify trainee performance by level better than Likert-type assessments. Average PGY-level scores from milestone-based assessments may ultimately provide guidance for determining whether trainees are progressing at the expected pace.
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.
Impact on house staff evaluation scores when changing from a Dreyfus- to a Milestone-based evaluation model: one internal medicine residency program's findings.


PURPOSE:
As graduate medical education (GME) moves into the Next Accreditation System (NAS), programs must take a critical look at their current models of evaluation and assess how well they align with reporting outcomes. Our objective was to assess the impact on house staff evaluation scores when transitioning from a Dreyfus-based model of evaluation to a Milestone-based model of evaluation. Milestones are a key component of the NAS.

METHOD:
We analyzed all end of rotation evaluations of house staff completed by faculty for academic years 2010-2011 (pre-Dreyfus model) and 2011-2012 (post-Milestone model) in one large university-based internal medicine residency training program. Main measures included change in PGY-level average score; slope, range, and separation of average scores across all six Accreditation Council for Graduate Medical Education (ACGME) competencies.

RESULTS:
Transitioning from a Dreyfus-based model to a Milestone-based model resulted in a larger separation in the scores between our three post-graduate year classes, a steeper progression of scores in the PGY-1 class, a wider use of the 5-point scale on our global end of rotation evaluation form, and a downward shift in the PGY-1 scores and an upward shift in the PGY-3 scores.

CONCLUSIONS:
For faculty trained in both models of assessment, the Milestone-based model had greater discriminatory ability as evidenced by the larger separation in the scores for all the classes, in particular the PGY-1 class.
Initial performance of a modified milestones global evaluation tool for semiannual evaluation of residents by faculty.


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.
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.
Positive Change in Feedback Perceptions and Behavior: A 10-Year Follow-up Study.


BACKGROUND:
Providing and learning from feedback are essential components of medical education, and typically described as resistant to change. But given a decade of change in the clinical context in which feedback occurs, the authors asked if, and how, perceptions of feedback and feedback behaviors might have changed in response to contextual affordances.

METHODS:
In 2017, the authors conducted a follow-up, ethnographic study on 2 general pediatric floors at the same children’s hospital where another ethnographic study on a general pediatric floor was conducted in 2007. Data sources included (1) 21 and 34 hours of observation in 2007 and 2017, respectively, (2) 35 and 25 interviews with general pediatric attending physicians and residents in 2007 and 2017, respectively, and (3) a review of 120 program documents spanning 2007 to 2017. Data were coded and organized around 3 recommendations for feedback that were derived from 2007 data and served as standards for assessing change in 2017.

RESULTS:
Data revealed progress in achieving each recommendation. Compared with 2007, participants in 2017 more clearly distinguished between feedback and evaluation; residents were more aware of in-the-moment feedback, and they had shifted their orientation from evaluation and grades to feedback and learning. Explanations for progress in achieving recommendations, which were derived from the data, pointed to institutional and national influences, namely, the pediatric milestones.

CONCLUSIONS:
On the basis of follow-up, ethnographic data, changes in the clinical context of pediatric education may afford positive change in perceptions of feedback and feedback behavior and point to influences within and beyond the institution.
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.
How do clinical competency committees use different sources of data to assess residents' performance on the internal medicine milestones? A mixed methods pilot study.


PURPOSE:
This study examines how Clinical Competency Committees (CCCs) synthesize assessment data to make judgments about residents' clinical performances.

METHODS:
Between 2014 and 2015, after four six-month reporting periods to the Accreditation Council for Graduate Medical Education (ACGME), 7 of 16 CCC faculty at Rush University Medical Center completed questionnaires focused on their perspectives about rating residents on their achievement of the milestones and participated in a focus group. Qualitative data were analyzed using grounded theory. Milestones ratings for two six-month ACGME reporting cycles (n = 100 categorical residents) were also analyzed.

RESULTS:
CCC members weighted resident rotation ratings highest (weight = 37%), followed by faculty rotation comments (weight = 27%) and personal experience with residents (weight = 14%) for making judgments about learner’s milestone levels. Three assessment issues were identified from qualitative analyses: (1) "design issues" (e.g. problems with available data or lack thereof); (2) "synthesis issues" (e.g. factors influencing ratings and decision-making processes) and (3) "impact issues" (e.g. how CCC generated milestones ratings are used).

CONCLUSIONS:
Identifying factors that affect assessment at all stages of the CCC process can contribute to improving assessment systems, including support for faculty development for CCCs. Recognizing challenges in synthesizing first and second-hand assessment data is an important step in understanding the CCC decision-making process.
Mapping Residency Global Health Experiences to the ACGME Family Medicine Milestones.


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.
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.
Using Milestones as Evaluation Metrics During an Emergency Medicine Clerkship.

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.
Development and evaluation of standardized narrative cases depicting the general surgery professionalism milestones.


PURPOSE:
Residency programs now are required to use educational milestones, which has led to the need for new methods of assessment. The literature suggests that narrative cases are a promising tool to track residents' progress. This study demonstrates the process for developing and evaluating narrative cases representing the five levels of the professionalism milestones.

METHOD:
In 2013, the authors identified 28 behaviors in the Accreditation Council for Graduate Medical Education general surgery professionalism milestones. They modified previously published narrative cases to fit these behaviors. To evaluate the quality of these cases, the authors developed a 28-item, five-point scale instrument, which 29 interdisciplinary faculty completed. The authors compared the faculty ratings by narrative case and specialty with the authors' initial rankings of the cases by milestone level. They used t tests and analysis of variance to compare mean scores across specialties.

RESULTS:
The authors developed 10 narrative cases, 2 for each of the 5 milestone levels. Each case contained at least 20 of the 28 behaviors identified in the milestones. Mean faculty ratings matched the milestone levels. Reliability was good (G coefficient = 0.86, phi coefficient = 0.85), indicating consistency in raters' ability to determine the proper milestone level for each case.

CONCLUSIONS:
The authors demonstrate a process for using specialty-specific milestones to develop narrative cases that map to a spectrum of professionalism behaviors. This process can be applied to other competencies and specialties to facilitate faculty awareness of resident performance descriptors and provide a frame of reference for milestones assessment.
The Milestones Passport: A Learner-Centered Application of the Milestone Framework to Prompt Real-Time Feedback in the Emergency Department


BACKGROUND:
In July 2013, emergency medicine residency programs implemented the Milestone assessment as part of the Next Accreditation System.

OBJECTIVE:
We hypothesized that applying the Milestone framework to real-time feedback in the emergency department (ED) could affect current feedback processes and culture. We describe the development and implementation of a Milestone-based, learner-centered intervention designed to prompt real-time feedback in the ED.

METHODS:
We developed and implemented the Milestones Passport, a feedback intervention incorporating subcompetencies, in our residency program in July 2013. Our primary outcomes were feasibility, including faculty and staff time and costs, number of documented feedback encounters in the first 2 months of implementation, and user-reported time required to complete the intervention. We also assessed learner and faculty acceptability.

RESULTS:
Development and implementation of the Milestones Passport required 10 hours of program coordinator time, 120 hours of software developer time, and 20 hours of faculty time. Twenty-eight residents and 34 faculty members generated 257 Milestones Passport feedback encounters. Most residents and faculty reported that the encounters required fewer than 5 minutes to complete, and 48% (12 of 25) of the residents and 68% (19 of 28) of faculty reported satisfaction with the Milestones Passport intervention. Faculty satisfaction with overall feedback in the ED improved after the intervention (93% versus 54%, P = .003), whereas resident satisfaction with feedback did not change significantly.

CONCLUSIONS:
The Milestones Passport feedback intervention was feasible and acceptable to users; however, learner satisfaction with the Milestone assessment in the ED was modest.
Operationalizing the internal medicine milestones—an early status report.

BACKGROUND:
The internal medicine milestones were developed to advance outcomes-based residency training and will play an important role in the next accreditation system.

INNOVATION:
As an element of our program’s participation in the internal medicine educational innovations project, we implemented a milestones-based evaluation process in our general medicine and pulmonary-critical care rotations on July 1, 2010.

MEASURES:
Outcomes assessed included survey-rated acceptability to participating faculty, residents, and clinical competency committee members.

RESULTS:
Faculty and residents agreed that the milestones promoted a common understanding of what knowledge, skills, and attitudes should be displayed at particular points in residents' professional development and enhanced evaluators' ability to provide specific performance feedback. Most residents and faculty members agreed that the milestones promoted fairness and uniformity in the evaluation process. Clinical competency committee members agreed that the milestones improved the quality of information available for deliberations and resulted in more uniform promotion standards. Faculty rated the use of too many milestones per form/tool at a mean of 7.3 (where 1 was minimally problematic, and 10 was maximally problematic) and the potential for evaluator fatigue (mean, 8.2) as the most significant challenges to the use of milestones. Eight of 12 faculty members would recommend milestones in other programs; 4 were uncertain.

CONCLUSIONS:
Despite logistical challenges, educators and trainees found that milestones promoted a common understanding of what knowledge, skills and attitudes should be displayed at particular stages of training; permitted greater specificity in performance feedback; and enhanced uniformity and fairness in promotion decisions.
Medical School Milestones Studies: Quantitative

Using Transitional Year Milestones to Assess Graduating Medical Students' Skills During a Capstone Course.


BACKGROUND:
Undergraduate medical education (UME) follows the lead of graduate medical education (GME) in moving to competency-based assessment. The means for and the timing of competency-based assessments in UME are unclear.

OBJECTIVE:
We explored the feasibility of using the Accreditation Council for Graduate Medical Education Transitional Year (TY) Milestones to assess student performance during a mandatory, fourth-year capstone course.

METHODS:
Our single institution, observational study involved 99 medical students who completed the course in the spring of 2014. Students' skills were assessed by self, peer, and faculty assessment for 6 existing course activities using the TY Milestones. Evaluation completion rates and mean scores were calculated.

RESULTS:
Students' mean milestone levels ranged between 2.2 and 3.6 (on a 5-level scoring rubric). Level 3 is the performance expected at the completion of a TY. Students performed highest in breaking bad news and developing a quality improvement project, and lowest in developing a learning plan, working in interdisciplinary teams, and stabilizing acutely ill patients. Evaluation completion rates were low for some evaluations, and precluded use of the data for assessing student performance in the capstone course. Students were less likely to complete separate online evaluations. Faculty were less likely to complete evaluations when activities did not include dedicated time for evaluations.

CONCLUSIONS:
Assessment of student competence on 9 TY Milestones during a capstone course was useful, but achieving acceptable evaluation completion rates was challenging. Modifications are necessary if milestone scores from a capstone are intended to be used as a handoff between UME and GME.
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.
Medical School Milestone Studies: Qualitative

Using the ACGME Milestones as a Handover Tool From Medical School to Surgery Residency.


OBJECTIVE:
To map current medical school assessments for graduating students to the Accreditation Council for Graduate Medical Education (ACGME) milestones in general surgery, and to pass forward individual performance metrics on level 1 milestones to receiving residency programs.

DESIGN:
The study included 20 senior medical students who were accepted into surgery internship positions. Data from medical school performance assessments from the third-year surgery clerkship, fourth-year surgery rotations, fourth-year surgery boot camp, Clinical Competency Assessment Examination, and United States Medical Licensing Examination (USMLE) Step 1 and 2 examinations were used to map each student's competency assessments to the General Surgery Milestones based on a scoring system created and validated by independent assessors. This Milestones Assessment was then provided to each student's receiving program director.

SETTING:
The study was conducted at the University of Michigan Medical School, in Ann Arbor, Michigan.

PARTICIPANTS:
Fourth-year medical students entering into surgical internship.

RESULTS:
Of 16 Accreditation Council for Graduate Medical Education (ACGME) General Surgery Milestones subcompetencies, 12 were able to be evaluated with current medical school assessments. Of the 20 students, 11 met criteria for all the level 1 milestones and 9 needed improvement in at least 1 domain.
PROBLEM:
Competency-based education, including assessment of specialty-specific milestones, has become the dominant medical education paradigm; however, how to determine baseline competency of entering interns is unclear—as is to whom this responsibility falls. Medical schools should take responsibility for providing residency programs with accurate, competency-based assessments of their graduates.

APPROACH:
A University of Michigan ad hoc committee developed (spring 2013) a post-Match, milestone-based medical student performance evaluation for seven students matched into emergency medicine (EM) residencies. The committee determined EM milestone levels for each student based on assessments from the EM clerkship, end-of-third-year multistation standardized patient exam, EM boot camp elective, and other medical school data.

OUTCOMES:
In this feasibility study, the committee assessed nearly all 23 EM milestones for all seven graduates, shared these performance evaluations with the program director (PD) where each student matched, and subsequently surveyed the PDs regarding this pilot. Of the five responding PDs, none reported using the traditional medical student performance evaluation to customize training, four (80%) indicated that the proposed assessment provided novel information, and 100% answered that the assessment would be useful for all incoming trainees.

NEXT STEPS:
An EM milestone-based, post-Match assessment that uses existing assessment data is feasible and may be effective for communicating competency-based information about medical school graduates to receiving residency programs. Next steps include further aligning assessments with competencies, determining the benefit of such an assessment for other specialties, and articulating the national need for an effective educational handover tool between undergraduate and graduate medical education institutions.
International Studies: Quantitative

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 $\alpha$, 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.
The R2C2 Model in Residency Education: How Does It Foster Coaching and Promote Feedback Use?


PURPOSE:
The authors previously developed and tested a reflective model for facilitating performance feedback for practice improvement, the R2C2 model. It consists of four phases: relationship building, exploring reactions, exploring content, and coaching. This research studied the use and effectiveness of the model across different residency programs and the factors that influenced its effectiveness and use.

METHOD:
From July 2014-October 2016, case study methodology was used to study R2C2 model use and the influence of context on use within and across five cases. Five residency programs (family medicine, psychiatry, internal medicine, surgery, and anesthesia) from three countries (Canada, the United States, and the Netherlands) were recruited. Data collection included audiotaped site assessment interviews, feedback sessions, and debriefing interviews with residents and supervisors, as well as completed learning change plans (LCPs). Content, thematic, template, and cross-case analysis were conducted.

RESULTS:
An average of 9 resident-supervisor dyads per site were recruited. The R2C2 feedback model, used with an LCP, was reported to be effective in engaging residents in a reflective, goal-oriented discussion about performance data, supporting coaching and enabling collaborative development of a change plan. Use varied across cases, influenced by six general factors: supervisor characteristics, resident characteristics, qualities of the resident-supervisor relationship, assessment approaches, program culture and context, and supports provided by the authors.

CONCLUSIONS:
The R2C2 model was reported to be effective in fostering a productive, reflective feedback conversation focused on resident development and in facilitating collaborative development of a change plan. Factors contributing to successful use were identified.
Commentaries/Editorials/Opinion Pieces/Other Milestone-related Studies:

Introducing a curriculum in ethics and professionalism for dermatology residencies.


Abstract
There is general agreement on what constitutes ethical reasoning and professional behavior, but standardized methods to teach these skills in dermatology residency are currently unavailable. We introduce a model curriculum designed to impart the knowledge and skills to meet the Accreditation Council for Graduate Medical Education Dermatology Milestones for Professionalism over a 3-year cycle.
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.
Advancing Simulation-Based Education in Pain Medicine.


BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) has recently implemented milestones and competencies as a framework for training fellows in Pain Medicine, but individual programs are left to create educational platforms and assessment tools that meet ACGME standards.

OBJECTIVES:
In this article, we discuss the concept of milestone-based competencies and the inherent challenges for implementation in pain medicine. We consider simulation-based education (SBE) as a potential tool for the field to meet ACGME goals through advancing novel learning opportunities, engaging in clinically relevant scenarios, and mastering technical and nontechnical skills.

RESULTS:
The sparse literature on SBE in pain medicine is highlighted, and we describe our pilot experience, which exemplifies a nascent effort that encountered early difficulties in implementing and refining an SBE program.

CONCLUSIONS:
The many complexities in offering a sophisticated simulated pain curriculum that is valid, reliable, feasible, and acceptable to learners and teachers may only be overcome with coordinated and collaborative efforts among pain medicine training programs and governing institutions.
Top Medical Education Studies of 2016: A Narrative Review.


Abstract

Education, like clinical medicine, should be based on the most current evidence in the field. Unfortunately, medical educators can be overwhelmed by the sheer volume and range of resources for this literature. This article provides an overview of 15 articles from 2016 that the authors consider the top articles in the field of pediatric medical education. The 7 authors, all medical educators with combined leadership and expertise across the continuum of pediatric medical education, used an iterative 3-stage process to review more than 6339 abstracts published in 2016. This process was designed to identify a small subset of articles that were most relevant to educational practices and most applicable to pediatric medical education. In the first 2 stages, pairs of authors independently reviewed and scored abstracts in 13 medical education-related journals and reached consensus to identify the articles that best met these criteria. In the final stage, all articles were discussed using a group consensus model to select the final articles included in this review. This article presents summaries of the 15 articles that were selected. The results revealed a cluster of studies related to observed standardized clinical encounters, self-assessment, professionalism, clinical teaching, competencies/milestones, and graduate medical education management strategies. We provide suggestions on how medical educators can apply the findings to their own practice and educational settings. This narrative review offers a useful tool for educators interested in keeping informed about the most relevant and valuable information in the field.
Mapping the Balint groups to the Accreditation Council for Graduate Medical Education family medicine competencies.


Abstract:
Introduction Balint group discussions provide learning opportunities for many of the competencies and milestones put forward by the Accreditation Council for Graduate Medical Education. The current literature is mixed concerning the effect of Balint groups on communication skills and professionalism. Aim To map the content of the Balint discussion to the competencies and milestones put forward by the Accreditation Council for Graduate Medical Education.

METHODOLOGY:
A group who were both experts in Balint and members of the clinical competency committee of residency programs rated narratives that summarized Balint group discussions. Credentialed Leaders of the American Balint Society were invited via email to submit narratives (250 words) about Balint groups that they have led, or were leading, with residents.

RESULTS:
Only four narratives were submitted. Additional cases were recruited through literature review of published Balint discussion cases. A total of 25 cases were rated by the committee. There was agreement between three out of four raters on at least one core milestone in every case. The most frequent milestones were C1 (develops meaningful therapeutic relationships with patients and families), C2 (communicated effectively with patients, families, and public), Prof1 (completes a process of professionalization), and Prof3 (demonstrates humanism and cultural proficiency). Balint groups provided a learning opportunity for a subset of milestones in at least 36% of the cases.

CONCLUSION:
This pilot research suggests that Balint groups and the discussions of complex and challenging cases provide learning opportunities for multiple family medicine milestones, mainly communication skills and professionalism. Further research is needed to refine the methodology and the rating system.
Bringing the Flipped Classroom to Day 1: A Novel Didactic Curriculum for Emergency Medicine Intern Orientation.


Abstract
Most emergency medicine (EM) residency programs provide an orientation program for their incoming interns, with the lecture being the most common education activity during this period. Our orientation program is designed to bridge the gap between undergraduate and graduate medical education by ensuring that all learners demonstrate competency on Level 1 Milestones, including medical knowledge (MK). To teach interns core medical knowledge in EM, we reformulated orientation using the flipped-classroom model by replacing lectures with small group, case-based discussions. Interns demonstrated improvement in medical knowledge through higher scores on a posttest. Evaluation survey results were also favorable for the flipped-classroom teaching format.
Competency-Based Medical Education and the Ghost of Kuhn: Reflections on the Messy and Meaningful Work of Transformation.


Abstract
The transition, if not transformation, to outcomes-based medical education likely represents a paradigm shift struggling to be realized. Paradigm shifts are messy and difficult but ultimately meaningful if done successfully. This struggle has engendered tension and disagreements, with many of these disagreements cast as either-or polarities. There is little disagreement, however, that the health care system is not effectively achieving the triple aim for all patients. Much of the tension and polarity revolve around how more effectively to prepare students and residents to work in and help change a complex health care system. Competencies were an initial attempt to facilitate this shift by creating frameworks of essential abilities needed by physicians. However, implementation of competencies has proven to be difficult. Entrustable professional activities (EPAs) in undergraduate and graduate medical education and Milestones in graduate medical education are recent concepts being tried and studied as approaches to guide the shift to outcomes. Their primary purpose is to help facilitate implementation of an outcomes-based approach by creating shared mental models of the competencies, which in turn can help to improve curricula and assessment. Understanding whether and how EPAs and Milestones effectively facilitate the shift to outcomes has been and will continue to be an iterative and ongoing reflective process across the entire medical education community using lessons from implementation and complexity science. In this Invited Commentary, the author reflects on what got the community to this point and some sources of tension involved in the struggle to move to outcomes-based education.
Are Milestones Really Measuring Development?


-no abstract available.
Competency milestones for medical students: Design, implementation, and analysis at one medical school.


Abstract
Competency-based assessment seeks to align measures of performance directly with desired learning outcomes based upon the needs of patients and the healthcare system. Recognizing that assessment methods profoundly influence student motivation and effort, it is critical to measure all desired aspects of performance throughout an individual's medical training. The Accreditation Council for Graduate Medical Education (ACGME) defined domains of competency for residency; the subsequent Milestones Project seeks to describe each learner's progress toward competence within each domain. Because the various clinical disciplines defined unique competencies and milestones within each domain, it is difficult for undergraduate medical education to adopt existing GME milestones language. This paper outlines the process undertaken by one medical school to design, implement and improve competency milestones for medical students. A team of assessment experts developed milestones for a set of focus competencies; these have now been monitored in medical students over two years. A unique digital dashboard enables individual, aggregate and longitudinal views of student progress by domain. Validation and continuous quality improvement cycles are based upon expert review, user feedback, and analysis of variation between students and between assessors. Experience to date indicates that milestone-based assessment has significant potential to guide the development of medical students.
Building a Framework of Entrustable Professional Activities, Supported by Competencies and Milestones, to Bridge the Educational Continuum.


Abstract
The transition to competency-based medical education (CBME) and adoption of the foundational domains of competence by the Accreditation Council for Graduate Medical Education, Association of American Medical Colleges (AAMC), and American Board of Medical Specialties' certification and maintenance of certification (MOC) programs provided an unprecedented opportunity for the pediatrics community to create a model of learning and assessment across the continuum. Two frameworks for assessment in CBME have been promoted: (1) entrustable professional activities (EPAs) and (2) milestones that define a developmental trajectory for individual competencies. EPAs are observable and measureable units of work that can be mapped to competencies and milestones critical to performing them safely and effectively. The pediatrics community integrated the two frameworks to create a potential pathway of learning and assessment across the continuum from undergraduate medical education (UME) to graduate medical education (GME) and from GME to practice. The authors briefly describe the evolution of the Pediatrics Milestone Project and the process for identifying EPAs for the specialty and subspecialties of pediatrics. The method of integrating EPAs with competencies and milestones through a mapping process is discussed, and an example is provided. The authors illustrate the alignment of the AAMC's Core EPAs for Entering Residency with the general pediatrics EPAs and, in turn, the alignment of the latter with the subspecialty EPAs, thus helping build the bridge between UME and GME. The authors propose how assessment in GME, based on EPAs and milestones, can guide MOC to complete the bridge across the education continuum.
The Promise of Milestones: Are They Living Up to Our Expectations?


-no abstract available.
Shifting Approaches for Evaluation of Resident Performance: From Competencies to Milestones. (Commentary)


-no abstract available.
Milestones and Competency-Based Medical Education in Internal Medicine.


-no abstract available.
Ensuring Resident Competence: A Narrative Review of the Literature on Group Decision Making to Inform the Work of Clinical Competency Committees.


Abstract
The expectation for graduate medical education programs to ensure that trainees are progressing toward competence for unsupervised practice prompted requirements for a committee to make decisions regarding residents' progress, termed a clinical competency committee (CCC). The literature on the composition of these committees and how they share information and render decisions can inform the work of CCCs by highlighting vulnerabilities and best practices. Objective We conducted a narrative review of the literature on group decision making that can help characterize the work of CCCs, including how they are populated and how they use information. Methods English language studies of group decision making in medical education, psychology, and organizational behavior were used. Results The results highlighted 2 major themes. Group member composition showcased the value placed on the complementarity of members' experience and lessons they had learned about performance review through their teaching and committee work. Group processes revealed strengths and limitations in groups' understanding of their work, leader role, and information-sharing procedures. Time pressure was a threat to the quality of group work. Conclusions Implications of the findings include the risks for committees that arise with homogeneous membership, limitations to available resident performance information, and processes that arise through experience rather than deriving from a well-articulated purpose of their work. Recommendations are presented to maximize the effectiveness of CCC processes, including their membership and access to, and interpretation of, information to yield evidence-based, well-reasoned judgments.
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.
Implementation of competency-based medical education: are we addressing the concerns and challenges?


CONTEXT:
Competency-based medical education (CBME) has emerged as a core strategy to educate and assess the next generation of physicians. Advantages of CBME include: a focus on outcomes and learner achievement; requirements for multifaceted assessment that embraces formative and summative approaches; support of a flexible, time-independent trajectory through the curriculum; and increased accountability to stakeholders with a shared set of expectations and a common language for education, assessment and regulation.

OBJECTIVES:
Despite the advantages of CBME, numerous concerns and challenges to the implementation of CBME frameworks have been described, including: increased administrative requirements; the need for faculty development; the lack of models for flexible curricula, and inconsistencies in terms and definitions. Additionally, there are concerns about reductionist approaches to assessment in CBME, lack of good assessments for some competencies, and whether CBME frameworks include domains of current importance. This study will outline these issues and discuss the responses of the medical education community.

METHODS:
The concerns and challenges expressed are primarily categorised as: (i) those related to practical, administrative and logistical challenges in implementing CBME frameworks, and (ii) those with more conceptual or theoretical bases. The responses of the education community to these issues are then summarised.

CONCLUSIONS:
The education community has begun to address the challenges involved in implementing CBME. Models and guidance exist to inform implementation strategies across the continuum of education, and focus on the more efficient use of resources and technology, and the use of milestones and entrustable professional activities-based frameworks. Inconsistencies in CBME definitions and frameworks remain a significant obstacle. Evolution in assessment approaches from in vitro task-based methods to in vivo integrated approaches is responsive to many of the theoretical and conceptual concerns about CBME, but much work remains to be done to bring rigour and quality to work-based assessment.
Fear, Regulations, and the Fragile Exoskeleton of Medical Professionalism.


-no abstract available.
Reflections on the First 2 Years of Milestone Implementation


BACKGROUND
The Accreditation Council for Graduate Medical Education (ACGME) and the American Board of Medical Specialties (ABMS) collectively constitute the foundation of professional self-regulation in the United States. In February 1999, the 2 organizations approved 6 general competencies broadly relevant for all medical practice, followed by the official launch of the Outcomes Project in 2001.1 It was expected that the competencies would be an antidote to overspecification of accreditation standards, and that they would empower programs to create training programs grounded in meaningful outcomes in a developmental approach.

As many programs can attest, the implementation of outcomes-based (eg, competency-based) medical education has been challenging. One reason has been the difficulty in implementing the competencies in both curriculum and assessment. Program leaders lacked shared mental models within their own training programs, accompanied by a lack of shared understanding nationally within disciplines. It is important to remember that one of the thorny problems the milestones were intended to address was the sources of unwanted and unwarranted variability in educational and, by extension, clinical outcomes. In addition, the community cannot improve at scale what cannot be measured, and prior frames and approaches to measurement were insufficient and ineffective. A key goal for milestones thus is to help improve the state and quality of measurement through better assessment in graduate medical education to facilitate the improved outcomes everyone desires.

Approximately 10 years ago, conversations began on how to more effectively and meaningfully operationalize the competencies to help improve the design of residency and fellowship programs through the use of a developmental framework. In parallel, the ACGME began to explore mechanisms to move the accreditation system to a focus on outcomes using a continuous quality improvement philosophy. Developmental milestones, using narratives to describe in more descriptive terms the professional trajectories of residents, were seen as a way to move the outcomes project forward. Starting in 2007, the disciplines of internal medicine, pediatrics, and surgery began to create developmental milestones for the 6 competencies.

Surgery would subsequently delay the development of their milestones focusing first on the SCORE curriculum. The ACGME began to restructure its accreditation processes in 2009, and soon after, milestone groups were constituted for all specialties. Milestone writing groups were cosponsored by the ACGME and the ABMS member certification boards. Early groups had significant latitude in developing their subcompetencies and milestones; specialties that started the process after 2010 used a standard template. Each milestone set was subjected to review by the educational community in the specialty. Box 1 provides an overview of the purposes of the milestones across key stakeholders, and figure 1 provides an example of a key driver diagram of milestones as an educational and clinical intervention. As figure 1 highlights, milestones can potentially trigger a number of drivers, or mechanisms, to help enable changes in residency and fellowship education.
Achieving the Desired Transformation: Thoughts on Next Steps for Outcomes-Based Medical Education.


Abstract
Since the introduction of the outcomes-based medical education (OBME) movement, progress toward implementation has been active but challenging. Much of the angst and criticism has been directed at the approaches to assessment that are associated with outcomes-based or competency frameworks, particularly defining the outcomes. In addition, these changes to graduate medical education (GME) are concomitant with major change in health care systems--specifically, changes to increase quality and safety while reducing cost. Every sector, from medical education to health care delivery and financing, is in the midst of substantial change and disruption. The recent release of the Institute of Medicine's report on the financing and governance of GME highlights the urgent need to accelerate the transformation of medical education. One source of continued tension within the medical education community arises from the assumption that the much-needed increases in value and improvement in health care can be achieved by holding the current educational structures and architecture of learning in place while concomitantly withdrawing resources. The authors of this Perspective seek to reframe the important and necessary debate surrounding the current challenges to implementing OBME. Building on recent change and service theories (e.g., Theory U and coproduction), they propose several areas of redirection, including reexamination of curricular models and greater involvement of learners, teachers, and regulators in cocreating new training models, to help facilitate the desired transformation in medical education.
Realizing the promise of competency-based medical education.


Abstract
Competency-based medical education (CBME) places a premium on both educational and clinical outcomes. The Milestones component of the Next Accreditation System represents a fundamental change in medical education in the United States and is part of the drive to realize the full promise of CBME. The Milestones framework provides a descriptive blueprint in each specialty to guide curriculum development and assessment practices. From the beginning of the Outcomes project in 1999, the Accreditation Council for Graduate Medical Education and the larger medical education community recognized the importance of improving their approach to assessment. Work-based assessments, which rely heavily on the observations and judgments of clinical faculty, are central to a competency-based approach. The direct observation of learners and the provision of robust feedback have always been recognized as critical components of medical education, but CBME systems further elevate their importance. Without effective and frequent direct observation, coaching, and feedback, the full potential of CBME and the Milestones cannot be achieved. Furthermore, simply using the Milestones as end-of-rotation evaluations to "check the box" to meet requirements undermines the intent of an outcomes-based accreditation system. In this Commentary, the author explores these challenges, addressing the concerns raised by Williams and colleagues in their Commentary. Meeting the assessment challenges of the Milestones will require a renewed commitment from institutions to meet the profession's "special obligations" to patients and learners. All stakeholders in graduate medical education must commit to a professional system of self-regulation to prepare highly competent physicians to fulfill this social contract.
Considering "Nonlinearity" Across the Continuum in Medical Education Assessment: Supporting Theory, Practice, and Future Research Directions.


Abstract
The purpose of this article is to propose new approaches to assessment that are grounded in educational theory and the concept of "nonlinearity." The new approaches take into account related phenomena such as "uncertainty," "ambiguity," and "chaos." To illustrate these approaches, we will use the example of assessment of clinical reasoning, although the principles we outline may apply equally well to assessment of other constructs in medical education. Theoretical perspectives include a discussion of script theory, assimilation theory, self-regulated learning theory, and situated cognition. Assessment examples to include script concordance testing, concept maps, self-regulated learning microanalytic technique, and work-based assessment, which parallel the above-stated theories, respectively, are also highlighted. We conclude with some practical suggestions for approaching nonlinearity.
The new milestones: do we need to take a step back to go a mile forward?


Abstract
The Milestones Project, like all previous systems and changes in graduate psychiatric education, for example, moving from 3 to 4 years of training or adopting six competency domains, has been devised without any supporting data and does not assess meaningful outcomes, such as improved patient outcomes. No evidence is presented that Milestones-based training will produce better psychiatrists. There is a path forward. First, replace unproven expert consensus with scientific and evidence-based approaches. Second, exchange endpoints that are easy to assess but uncorrelated with real world functioning (e.g., multiple-choice examinations) for outcomes that are meaningful and external to the training program (e.g., patient outcomes). Finally, to prevent possible waste, excess burden, or harm, no changes should be mandated until proven in prospective studies.
Building the continuum of competency-based medical education.


Paul Batalden is credited with the observation that, ‘every system is perfectly designed for the results it generates’. In this volume of Perspectives in Medical Education, Chen and associates highlight the importance of this observation by describing self-reported gaps in preparedness of learners transitioning from undergraduate to graduate medical training. By identifying these gaps, these authors highlight one of the most serious challenges facing the medical education community as it operationalizes competency-based medical education. Using Batalden’s axiom, learners experience a medical education system that is perfectly designed to inadequately prepare them for the next stage of their professional development and ultimately to work competently in the health care delivery system of the future.
Two Cheers for Milestones.


This editorial will explore the implementation of milestones across graduate medical education (GME) from 2 perspectives. The first is my perspective as a clinician, who often asks, “How do I make decisions with a patient when there isn’t evidence to use as a guideline?” The second is my perspective as a department chair who asks a different question: “What resources are needed for milestone implementation?”
Medical specialty boards can help measure graduate medical education outcomes.


Abstract
U.S. graduate medical education (GME) training institutions are under increasing scrutiny to measure program outcomes as a demonstration of accountability for the sizeable funding they receive from the federal government. The Accreditation Council for Graduate Medical Education (ACGME) is a potential agent of measuring GME accountability but has no interaction with physicians after residency training is completed. American Board of Medical Specialty (ABMS) member boards interact with physicians throughout their careers through maintenance of certification (MOC) and are a potential source of valuable data on physician competency and quality of care, both of which could be used to measure GME accountability. The authors propose that ABMS boards and the ACGME deepen their existing relationship to better assess residency training outcomes. ABMS boards have a wealth of data on physicians collected as a by-product of MOC and business operations. Further, many ABMS boards collect practice demographics and scope-of-practice information through MOC enrollment surveys or recertification examination questionnaires. These data are potentially valuable in helping residencies know what their graduates are doing in practice. Part 4 of MOC generally involves assessment of the quality of care delivered in practice, and ABMS boards could share these deidentified data with the ACGME and residency programs to provide direct feedback on the practice outcomes of graduates. ABMS member boards and the ACGME should broaden their long-standing relationship to further develop shared roles and data-sharing mechanisms to better inform residencies and the public about GME training outcomes.
Time to trust: longitudinal integrated clerkships and entrustable professional activities.


Abstract
Medical education shaped by the learning sciences can better serve medical students, residents, faculty, health care institutions, and patients. With increasing innovation in undergraduate and graduate medical education and more focused attention on educational principles and how people learn, this era of educational transformation offers promise. Principles manifest in "educational continuity" are informing changes in educational structures and venues and are enriching new discourse in educational pedagogy, assessment, and scholarship. The articles by Myhre and colleagues and Woloschuk and colleagues in this issue, along with mounting evidence preceding these works, should reassure that principle-driven innovation in medical education is not only possible but can be achieved safely. In this commentary, the authors draw from these works and the wider literature on longitudinal integrated educational design. They suggest that the confluences of movements for longitudinal integrated clerkships and entrustable professional activities open new possibilities for other educational and practice advancements in quality and safety. With the advent of competency-based education, explicit milestones, and improved assessment regimens, overseers will increasingly evaluate students, trainees, and other learners on their ability rather than relying solely on time spent in an activity. The authors suggest that, for such oversight to have the most value, assessors and learners need adequate oversight time, and redesign of educational models will serve this operational imperative. As education leaders are reassessing old medical school and training models, rotational blocks, and other barriers to progress, the authors explore the dynamic interplay between longitudinal integrated learning models and entrustment.
Applying the milestones in an internal medicine residency program curriculum: a foundation for outcomes-based learner assessment under the next accreditation system.


Abstract
In 2010, University of Kansas Medical Center internal medicine residency program leaders concluded that their competency-based curriculum and evaluation system was not sufficient to promote accurate assessment of learners' performance and needed revision to meet the requirements of the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System (NAS). Evaluations of learners seldom referenced existing curricular goals and objectives and reflected an "everyone is exceptional, no one is satisfactory" view. The authors identified the American Board of Internal Medicine and ACGME's Developmental Milestones for Internal Medicine Residency Training as a published standard for resident development. They incorporated the milestones into templates, a format that could be modified for individual rotations. A milestones-based curriculum for each postgraduate year of training and every rotation was then created, with input from educational leaders within each division in the Department of Internal Medicine and with the support of the graduate medical education office. In this article, the authors share their implementation process, which took approximately one year, and discuss their current work to create a documentation system for direct observation of entrustable professional activities, with the aim of providing guidance to other programs challenged with developing an outcomes-based curriculum and assessment system within the time frame of the NAS.
Milestones for Apheresis education.


Abstract
Milestones represent the essential knowledge, skills, and attitudes required for the practice of a medical discipline. Defining these milestones for each medical specialty has become a focus for the American Council of Graduate Medical Education (ACGME). Practitioners of Apheresis Medicine come from a variety of medical specialties making it challenging to establish the essential educational milestones for all. The American Society for Apheresis (ASFA) has an interest in promoting standards of excellence for Apheresis Medicine. ASFA’s Physician’s Curriculum Content Committee is a group of physician educators in the field of Apheresis Medicine, both donor and therapeutic, from across the United States, who have met regularly for several years to discuss the appropriate educational milestones in Apheresis training. The committee members teach residents and fellows from Pathology, Transfusion Medicine, Hematology/Oncology, Nephrology and other specialties. In this document, we have outlined the basic set of Apheresis milestones required in the ACGME defined competency areas of Patient Care and Medical Knowledge. We have also recommended methods of evaluation and estimated the time necessary for the acquisition of these cognitive and behavioral elements.
The next GME accreditation system—rationale and benefits.


-no abstract available.
Appendix:

Original Description of the Milestones by Specialty (*Journal of Graduate Medical Education (JGME)* supplementary issue, March 2013)

**A Milestone in the Milestones Movement: the JGME Milestones Supplement**
Gail Sullivan, Deborah Simpson, Thomas Cooney, and Eugene Beresin
Journal of Graduate Medical Education 2013 5:1s1, 1-4

**Emergency Medicine Milestones**
Journal of Graduate Medical Education 2013 5:1s1, 5-13

**Internal Medicine Milestones**
William Iobst, Eve Aagaard, Hasan Bazari, Timothy Brigham, Roger W. Bush, Kelly Caverzagie, Davoren Chick, Michael Green, Kevin Hinchey, Eric Holmboe, Sarah Hood, Gregory Kane, Lynne Kirk, Lauren Meade, Cynthia Smith, and Susan Swing
Journal of Graduate Medical Education 2013 5:1s1, 14-23

**Neurological Surgery Milestones**
Journal of Graduate Medical Education 2013 5:1s1, 24-35

**Orthopaedic Surgery Milestones**
Journal of Graduate Medical Education 2013 5:1s1, 36-58

**Pediatrics Milestones**
Carol Carraccio, Bradley Benson, Ann Burke, Robert Engleander, Susan Guralnick, Patricia Hicks, Stephen Ludwig, Daniel Schumacher, and Jerry Vasilias
Journal of Graduate Medical Education 2013 5:1s1, 59-73

**Diagnostic Radiology Milestones**
Journal of Graduate Medical Education 2013 5:1s1, 74-78

**Urology Milestones**
Michael Coburn, Christopher Amling, Robert R. Bahnson, Philipp Dahm, B. Price Kerfoot, Louise King, Brian Lane, Michael L. Ritchey, Charles D. Scales, Jr, Chandru P. Sundaram, and Susan R. Swing
Journal of Graduate Medical Education 2013 5:1s1, 79-98
Original Description of the Milestones by Specialty (*Journal of Graduate Medical Education (JGME)* supplementary issue, March 2014)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Authors</th>
<th>Journal and Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulating the Allergy and Immunology Milestones</td>
<td>Michael R. Nelson</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 1-4</td>
</tr>
<tr>
<td>The Allergy and Immunology Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 5-11</td>
</tr>
<tr>
<td>Development of the Anesthesiology Educational Milestones</td>
<td>Scott A. Schartel, Catherine Kuhn, Deborah J. Culley, Margaret Wood, and Neal Cohen</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 12-14</td>
</tr>
<tr>
<td>The Anesthesiology Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 15-28</td>
</tr>
<tr>
<td>The Development of the Colon and Rectal Surgery Milestones</td>
<td>Charles B. Whitlow</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 29-30</td>
</tr>
<tr>
<td>The Colon and Rectal Surgery Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 31-43</td>
</tr>
<tr>
<td>The Dermatology Milestone Project</td>
<td>George W. Turiansky and Daniel Loo</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 47-70</td>
</tr>
<tr>
<td>The Dermatology Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 44-46</td>
</tr>
<tr>
<td>Development of the Family Medicine Milestones</td>
<td>Suzanne Allen</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 71-73</td>
</tr>
<tr>
<td>The Family Medicine Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 74-86</td>
</tr>
<tr>
<td>The Medical Genetics Residency Milestones</td>
<td>Bruce R. Korf</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 87-90</td>
</tr>
<tr>
<td>The Medical Genetics Milestone Project</td>
<td></td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 91-101</td>
</tr>
<tr>
<td>Introducing the Neurology Milestones</td>
<td>Steven L. Lewis, Ralph F. Józefowicz, Shannon Kilgore, Amar Dhand, and Laura Edgar</td>
<td><em>Journal of Graduate Medical Education</em> 2014 6:1s1, 102-104</td>
</tr>
</tbody>
</table>
The Neurology Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 105-115

Nuclear Medicine Milestones
Lorraine M. Fig and Laura Edgar
Journal of Graduate Medical Education 2014 6:1s1, 116-118

The Nuclear Medicine Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 119-125

Obstetrics and Gynecology Milestones
Jessica L. Bienstock, Laura Edgar, and Rebecca McAlister
Journal of Graduate Medical Education 2014 6:1s1, 126-128

The Obstetrics and Gynecology Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 129-143

Developing the Educational Milestones for Ophthalmology
Anthony Arnold
Journal of Graduate Medical Education 2014 6:1s1, 144-145

The Ophthalmology Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 146-161

Developing the Otolaryngology Milestones
Terance T. Tsue
Journal of Graduate Medical Education 2014 6:1s1, 162-165

The Otolaryngology Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 166-179

Pathology Milestones
Wesley Y. Naritoku and C. Bruce Alexander
Journal of Graduate Medical Education 2014 6:1s1, 180-181

The Pathology Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 182-203

The Development of the Physical Medicine and Rehabilitation Milestones
William L. Bockenek, Teresa L. Massagli, Susan R. Swing, and Caroline Fischer
Journal of Graduate Medical Education 2014 6:1s1, 204-206

The Physical Medicine and Rehabilitation Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 207-221

The Plastic Surgery Milestone Project
Mary H. McGrath
Journal of Graduate Medical Education 2014 6:1s1, 222-224

The Plastic Surgery Milestone Project
Journal of Graduate Medical Education 2014 6:1s1, 225-242
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Preventive Medicine Milestone Project</strong></td>
<td>Sandra G. Delgado</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 243-245</td>
<td></td>
</tr>
<tr>
<td><strong>The Preventive Medicine Milestone Project: Aerospace Medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The Preventive Medicine Milestone Project: Occupational Medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The Preventive Medicine Milestone Project: Public Health and General Preventive Medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Educational Milestone Development for Psychiatry</strong></td>
<td>Christopher R. Thomas</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 281-283</td>
<td></td>
</tr>
<tr>
<td><strong>The Psychiatry Milestone Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>The Development of the Radiation Oncology Milestones</strong></td>
<td>W. Robert Lee and Robert J. Amdur</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 305-306</td>
<td></td>
</tr>
<tr>
<td><strong>The Radiation Oncology Milestone Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Development of the Educational Milestones for Surgery</strong></td>
<td>Thomas H. Cogbill and Susan R. Swing</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 317-319</td>
<td></td>
</tr>
<tr>
<td><strong>The General Surgery Milestone Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Educational Milestone Development in Phase II Specialties: Thoracic Surgery</strong></td>
<td>Stephen C. Yang and Walter Merrill</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 329-331</td>
<td></td>
</tr>
<tr>
<td><strong>The Thoracic Surgery Milestone Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Educational Milestone Development for Transitional Year Residency Training</strong></td>
<td>Steven R. Craig and Danny M. Takanishi, Jr</td>
<td>Journal of Graduate Medical Education 2014 6:1s1, 355-357</td>
<td></td>
</tr>
<tr>
<td><strong>The Transitional Year Milestone Project</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>