# Impact of an entrustable professional activities-based assessment system

Megan Lord, MD,[1] Amanda B. Murchison, MD,[2] Alaina Lawrence, BS,[2] Isaiah Johnson, MD[2]

**Keywords**: Resident evaluations, entrustable professional activities, milestones

***Abstract***

*Background: Beginning in 2014, all Accreditation Counsel of Graduate Medical Education (ACGME) accredited residency programs were required to move to a Milestones-based system for biannual resident assessment. The resident assessment system for the Virginia Tech-Carilion Obstetrics and Gynecology (OB/GYN) residency program was re-designed to meet this requirement in July, 2014. The ACGME Milestones based assessment tool was identified on multiple faculty surveys as an area for improvement. To address this issue, an entrustable professional activities (EPA) based assessment system was designed and implemented for assessment of all OB/GYN rotations.*

*Objective: To evaluate the impact of an EPA based resident assessment system on faculty member's evaluation of resident assessment tools.*

*Methods: In this prospective quality improvement study, a survey was sent to all faculty members prior to the implementation of the EPA-based assessment system. The same survey was performed three months after the implementation of the new system. To facilitate analysis, each level of agreement was assigned a numerical value (1-5). The results were aggregated, and were analyzed using t-tests, assuming unequal variances.*

*Results: Sixty-eight percent of the faculty responded to the first survey, and 67% responded to the follow up survey. Statistically significant ($p<.05$) improvements were noted in most measures of the EPA based assessment tool including "ease of use" (2.2 vs 4.4, $p< 0.001$) and "accurate representation of resident performance" (2.5 vs 3.9, $p <0.001$).*

*Conclusion: An EPA based resident evaluation system significantly improved teaching faculty's impression of most domains of our OB/GYN resident assessment tools.*

*[1]Carilion Clinic, Virginia Tech Carilion School of Medicine, Roanoke, Virginia*
*[2]Department of Obstetrics and Gynecology, Virginia Tech Carilion School of Medicine, Salem, Virginia*

*Please cite this paper as: Lord M, Amanda B. Murchison AB, Lawrence A, Johnson I. Impact of an entrustable professional activities-based assessment system. Proc Obstet Gynecol. 2018;8(3):Article 28 [ 9 p.]. Available from: http://ir.uiowa.edu/pog/. Free full text article.*

***Corresponding author:*** *Amanda Murchison, MD, Residency Program Director, Associate Professor, Department of Obstetrics and Gynecology, Virginia Tech Carilion School of Medicine, Salem, Virginia. abmurchison@carilionclinic.org*

***Financial Disclosure****: The authors report no conflict of interest.*

## Introduction

Beginning in 2014, all Accreditation Council of Graduate Medical Education (ACGME) accredited residency programs were required to report the progress of each resident using a Milestones-based system biannually. The Virginia Tech-Carilion Obstetrics and Gynecology (OB/GYN) residency program, like many other programs across the country, re-designed its resident assessment tool to address this requirement in July, 2014. The new assessment system directly incorporated relevant Milestones into the assessment forms. The ACGME Milestones based tool was identified on multiple faculty surveys as an area for improvement. To address this issue, the format of the assessment was changed in July 2016 to a system based on entrustable professional activities (EPA).

EPAs represent a shift in resident assessment. Unlike general competencies, EPAs focus on clinical outcomes and activities. This allows faculty to assess their learners using broad domains that include knowledge, skills, professionalism and communication all within a single EPA.[1] Previous research has shown that ad hoc entrustment decisions in the clinical setting are made by clinical supervisors primarily based on each resident's experience and exposure to a particular task, rather than generic cross-context competencies such as communication skills.[2] EPAs integrate competencies, which are not observable in isolation, into observable professional tasks, allowing direct assessment of the resident's work performance rather than attempting to infer competence based on underlying generic competencies. EPAs acknowledge how performance is situation- and task-dependent.[3] EPAs represent the important, observable activities that must be mastered before a resident is ready for safe, independent practice within a given specialty. They describe the work, communication and professionalism you would expect from a competent physician. When EPAs are carefully designed, they can define a particular specialty; while at the same time, set expectations for learners.[4,5,6] Assessing medical residents using EPAs relies on experienced faculty to make entrustment decisions.[1] These assessments of trustworthiness take into account more than just an assessment of clinical skills, EPAs require evaluators to consider multiple dimension of the learners performance during daily activities: knowledge, clinical skills, discernment, conscientiousness and truthfulness.[7] To that end, the rating scale does not include numeric levels which might mentally translate to a postgraduate year or to a "grade," but instead describes where the learner is on the continuum to independent practice for each task (Figure 1.) Within the assessment system, each EPA is "mapped" to a number of related Milestones. Representative EPAs and their related Milestones can be seen in Table 1. Within the new assessment system, the score given on each EPA is translated, without any additional input from the rater, to a rating on each of these linked Milestones, which are compiled for the Clinical Competency Committee to periodically review. Linking EPAs to Milestones does have limitations. Each EPA has multiple domains and each Milestone level has 1-5 sub-components. By linking,

assumptions are made that the resident is performing equally well in each domain as well as each Milestone task within a particular subsection.

The purpose of this study was to compare the acceptability and ease of use as determined by teaching faculty for an EPA based assessment system as compared with the current Milestones based evaluation system.

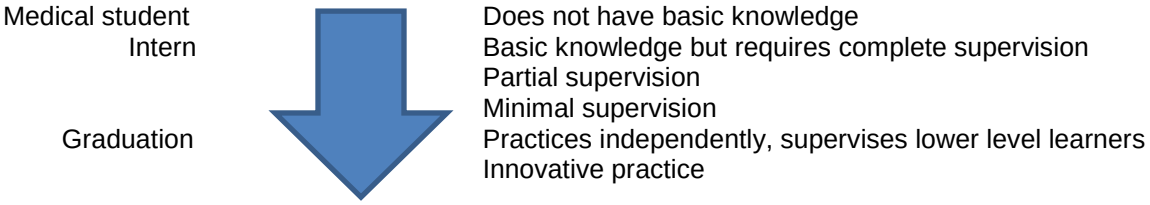


**Figure 1: Rating Scale for EPA-based Assessment**

**Table 1. Sample Mapped Entrustable Professional Activities to the Obstetrics and Gynecology Milestones**

| EPA | MK | PC | ICS | PROF | PBLI | SBP |
|---|---|---|---|---|---|---|
| Develops effective diagnostic and management plans for patients with first trimester bleeding and individualizes the management plan using best available evidence | MK6 | | ISC1 | PROF3 | PBLI1 | |
| Evaluate patients presenting with pelvic masses and develop appropriate management plans utilizing a rational and cost-effective approach to imaging, laboratory testing and surgical intervention | MK4 | | ICS3 | | PBLI1 | |
| Manage the perioperative care of low- and high-risk patients undergoing gynecologic surgery | MK1 | | | | | SBP2 |
| Recognizes and manages obstetrical emergencies | | PC1<br>PC2<br>PC3 | ICS1 | | | |
| MK= Medical Knowledge PC= Patient Care ICS= Interpersonal Communication and Skills<br>PBLI= Problem Based Learning & Improvement SBP= Systems Based Practice | | | | | | |

## Methods

*EPA Development*

EPAs for each practice setting were developed by the faculty members who oversee the residents in that setting using a process modified from the work of Kwan, et al.[8] Members of the OB/GYN Program Evaluation Committee were instructed on how to write an EPA using ten Cates guidelines which he described in his article entitled "Nuts and Bolts of Entrustable Professional Activities."[3] Following this, the Program Evaluation Committee members met with their respective divisions to develop a list of EPAs which comprised the "essential work" of the rotation. These lists were reviewed by

the Program Director and Assistant Program Director and presented to the full Program Evaluation Committee for approval. 70 EPAs were approved and mapped to the applicable Milestones. These included 2-16 EPAs for each of the following rotations: obstetrics, gynecology, maternal fetal medicine, reproductive endocrinology and fertility, urogynecology, pediatric adolescent gynecology and gynecologic oncology. Each EPA was mapped to the applicable Milestones and rotation specific evaluations were created in the MedHub platform (MedHub, Minneapolis, MN). A sample of the MedHub evaluation for the obstetrics rotation can be found in Table 2.

**Table 2. Sample EPA Assessment for the Obstetrics Rotation**

| EPA | Does not have basic knowledge | Basic knowledge, but requires complete supervision | Partial supervision | Minimal supervision | Practices independently, supervises lower level learners | Innovative practice |
|---|---|---|---|---|---|---|
| **Demonstrates a comprehensive understanding of the presentation of medical and obstetrical complications of pregnancy, appropriately counsels patients and makes cost effective management plans, including consults when appropriate** | | | | | | |
| **Provides care and communicates plans for women with abnormal labor or complex intrapartum conditions.** | | | | | | |
| **Performs complicated vaginal deliveries/ operative vaginal deliveries independently.** | | | | | | |
| **Appropriately consents patient for and performs a complicated cesarean.** | | | | | | |
| **Recognizes, repairs and manages obstetrical lacerations.** | | | | | | |
| **Recognizes and** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **manages obstetrical emergencies.** | | | | | | |
| **Provides postpartum care and directed counseling; recognizing and managing complications while considering cost and socioeconomic barriers.** | | | | | | |
| **Manages the obstetrical service, teaching others and serving as a consultant.** | | | | | | |
| **Consistently models compassion, integrity and respect for others, navigates ethically complex situations and coaches others.** | | | | | | |

*Survey Methods*

All department teaching faculty members received an anonymous, web-based survey to assess the Milestones-based evaluation system. This survey asked respondents to use a 5-point Likert scale to rate their level of agreement or disagreement with seven discrete statements (Table 3) about the resident evaluation tool (disagree, somewhat disagree, neutral, somewhat agree, agree). Free-text comment boxes were provided for both strengths and weaknesses of the evaluation tool. If 20% or more of respondents had similar comments then this would be identified as a recurring "theme". It was thought that these themes may be important if future modifications were needed. The identical survey was distributed to the teaching faculty 12 weeks after introduction of the new EPA-based assessment tool.

Each level of agreement (disagree to agree) was assigned a numerical value from one to five, with disagree receiving a score of one, and agree receiving a score of five. The responses were aggregated, and mean agreement scores were calculated. Mean agreement scores were compared using a t-test, assuming unequal variances. P-values of less than 0.05 were considered significant.

**Table 3. Results**

| Item | Milestone Evaluation (Mean) | EPA Evaluation (Mean) | P-value |
|---|---|---|---|
| **The current evaluation tool is easy to use** | 2.2 | 4.4 | 0.0001 |
| **I understand the rating scale of the current evaluation tool*** | 4.2 | 4.7 | .0743 |
| **The current evaluation tool results in useful feedback to the residents** | 2.3 | 3.9 | .0013 |
| **Completing the evaluation tool fits easily into my workday** | 1.9 | 3.9 | .00002 |
| **The current evaluation tool allows an accurate representation of residents' performance.** | 2.5 | 3.9 | 0.0007 |
| **The current evaluation tool provides consistent inter-observer rating of individual residents** | 2.4 | 3.8 | .0011 |
| **I am satisfied with the current resident evaluation tool** | **1.6** | **4.0** | **.000001** |

*not statistically significant

Each level of agreement (disagree to agree) was assigned a numerical value from one to five, with disagree receiving a score of one, and agree receiving a score of five. The responses were aggregated, and mean agreement scores were calculated. Mean agreement scores were compared using a t-test, assuming unequal variances. P-values of less than 0.05 were considered significant.

## Results

15 of 22 faculty members (68%) responded to the initial survey, and 14 of 21 faculty members (67%) responded to the follow-up survey. One faculty member left the department between the first and second survey, and was therefore not sent the second survey to complete. Results of T-test analyses of the faculty responses are summarized in Table 3. Overall satisfaction with the evaluation tool rose notably, with 100% of respondents in the initial survey either dissatisfied or neutral with the Milestones-based tool overall, compared to 7% dissatisfied, 21% neutral, and 71% satisfied with the new tool (mean agreement 4.0 vs 1.6, $p<0.001$. Respondents overwhelmingly agreed that the EPA-based evaluation tool is easier to use than the Milestones-based tool (mean agreement score 4.4 vs 2.2, p,0.001), that the new tool fits more easily into their workday (3.9 vs 1.9, $p<0.001$), and that the new tool provides a more accurate representation of residents' performance (3.9 vs 2.5, $p<0.001$).

The majority of respondents also left comments on the strengths of both evaluation tools. Representative comments are shown below in Table 4. Common themes included that the tasks described in the Milestones are not commonly encountered by the evaluators, and that the Milestones-based evaluations required a large amount of time to complete. Several (3 out of 15 respondents) of the faculty described the Milestones-based evaluation tool as "cumbersome." With regard to positive aspects of the Milestones-based tool, 3/15 respondents (20%) identified the

comment box as a strength. Respondents felt that the EPA-based tool was more concise, more efficient, and easier to complete. Weaknesses identified appeared to be more general weaknesses of an online evaluation tool; that, while the tool is much easier to use, feedback should still be reviewed with residents face-to-face.

**Table 4. Representative Comments**

| **Milestones-Based Tool** | |
|---|---|
| **Strengths** | **Weaknesses** |
| "It is standard across the board."<br><br>"They are all rated using the same scale and criteria"<br><br>"Easily maps to milestones [because] they are the milestones and milestones must be logged for each resident twice a year"<br><br>"The comment box"<br><br>"More descriptive, less punitive" | "Most of the categories as currently written are not applicable to my subspecialty"<br><br>"Evaluation categories and/or milestones don't equate with typical resident training experience and achievements during the rotations."<br><br>"Time consuming to complete"<br><br>"Cumbersome to use, unsure if it conveys appropriate feedback to the resident"<br><br>"Does not really address objective competency" |
| **EPA-Based Tool** | |
| **Strengths** | **Weaknesses** |
| "Updated version is more efficient and appropriate"<br><br>"Easy to complete and focused"<br><br>"Lack of redundancy and ease of use."<br><br>"brevity" | "Need to encourage faculty to personally review feedback with the residents in real-time"<br><br>"May need still significant face to face feedback for the resident to find it useful." |

## Discussion

Milestone based assessments of residents was viewed by the faculty as cumbersome, non-intuitive and difficult to apply to clinical supervision. Entrustable professional activities link competencies to clinical activities; making the assessment tool more intuitive for faculty. Our results demonstrate that academic OB/GYN faculty rated evaluations based on EPAs as easier to complete, providing a more accurate representation of resident performance, and more helpful to residents than frequently rating residents directly on the Milestones.

With regard to the one measure that did not show a statistically significant difference, the ease of understanding the rating scale, this is likely because both ratings scales were overall felt to be easy to understand. The vast majority of respondents to each survey somewhat agreed or agreed that they understood the rating scale of the tool being evaluated (87% for the

milestones-based tool vs 93% for the EPA-based tool, p=0.156), demonstrating that the EPA-based rating scale is no more difficult to understand than a Milestones-based scale.

The EPA evaluation allows for faculty to assess residents on the work they have observed, creating a more intuitive evaluation. This evaluation is then translated through a mapping process to the appropriate Milestones. The advantage of this is that a significant amount of data is generated on each Milestone, without the reviewer needing to rate each Milestone specifically. Some subtlety is, of course, lost, as a single more global rating is "split out" and applied to multiple interrelated but not identical competencies.

We believe that the strengths of this study include a high response rate (67-68%) in both initial and follow-up surveys, and prospective nature of the study. Limitations of the study include a relatively small study group, which consisted of 21-22 faculty members at a single academic institution. Further studies should be done to replicate these findings at larger institutions and in other specialties. This study also only assessed faculty perceptions of the new evaluation tool – it did not assess residents' perceptions of the new tool, or impact of the new system on the work of the Clinical Competency Committee. In addition, only aggregate data from the survey was collected for analysis; therefore, an intra class correlation coefficient could not be calculated. Several members of the faculty participated in the creation of the EPAs for each rotation; their participation in this project may have created bias in favor of the new system. Finally, we are making an assumption that you can translate an EPA assessment into a Milestone rating for the purpose of ACGME reporting. As mentioned previously, there are some limitations in the mapping process.

## Conclusions

The ACGME Milestones were intended to provide a framework to help monitor the developmental progress of each resident. The Milestones serve as a periodic summative feedback tool, to supplement each program's existing formative feedback tools. Instead, many programs, ours included, replaced their existing formative feedback tools with the Milestones. ACGME Milestones, while a useful summative feedback tool for evaluating competencies which span many professional activities and contexts, may not be the most intuitive tool for frontline faculty who we rely on to make assessments of residents. The new EPA-based evaluation tool was overwhelmingly viewed more favorably than the previous, Milestone-based evaluation tool. While developing EPAs for each specialty is a time-consuming, multi-step process, it appears that this initial investment pays off in faculty satisfaction and ease of evaluations later on. Once EPAs are developed, they can be mapped to Milestones such that Milestone data is still being continually collected, and frequent Milestone data remains available to the Clinical Competency Committee for review without requiring faculty members to rate residents on the Milestones directly at such frequent intervals.

*Acknowledgments: The authors wish to acknowledge Ellen Lockhart, MS, Carilion Clinic Biostatistician, who contributed statistical analysis services to this project.*

## References

1. Patel M, Baker P. Supervision for entrustable professional activities. Med Educ. 2018 Oct;52(10):998-1000. https://doi.org/10.1111/medu.13685 PubMed PMID: 30255530.

2. Van Loon KA, Teunissen PW, Driessen EW, Scheele F. The Role of Generic Competencies in the Entrustment of Professional Activities: A Nationwide Competency-Based Curriculum Assessed. J Grad Med Educ. 2016 Oct;8(4):546-552. https://doi.org/10.4300/JGME-D-15-00321.1 PubMed PMID: 27777665; PubMed Central PMCID: PMC5058587.

3. Ten Cate O. Nuts and bolts of entrustable professional activities. J Grad Med Educ. 2013 Mar;5(1):157-8. https://doi.org/10.4300/JGME-D-12-00380.1 PubMed PMID: 24404246; PubMed Central PMCID: PMC3613304.

4. Mink RB, Myers AL, Turner DA, Carraccio CL. Competencies, Milestones, and a Level of Supervision Scale for Entrustable Professional Activities for Scholarship. Acad Med. 2018 Nov;93(11):1668-1672. https://doi.org/10.1097/ACM.0000000000002353 PubMed PMID: 29995669.

5. Ten Cate O. A primer on entrustable professional activities. Korean J Med Educ. 2018 Mar;30(1):1-10. https://doi.org/10.3946/kjme.2018.76 Epub 2018 Feb 28. PubMed PMID: 29510603; PubMed Central PMCID: PMC5840559.

6. Carraccio C, Englander R, Gilhooly J, Mink R, Hofkosh D, Barone MA, Holmboe ES. Building a Framework of Entrustable Professional Activities, Supported by Competencies and Milestones, to Bridge the Educational Continuum. Acad Med. 2017 Mar;92(3):324-330. https://doi.org/10.1097/ACM.0000000000001141 PubMed PMID: 26959225.

7. Kennedy TJ, Regehr G, Baker GR, Lingard L. Point-of-care assessment of medical trainee competence for independent clinical work. Acad Med. 2008 Oct;83(10 Suppl):S89-92. https://doi.org/10.1097/ACM.0b013e318183c8b7 PubMed PMID: 18820510.

8. Kwan J, Crampton R, Mogensen LL, Weaver R, van der Vleuten CP, Hu WC. Bridging the gap: a five stage approach for developing specialty-specific entrustable professional activities. BMC Med Educ. 2016 Apr 20;16:117. https://doi.org/10.1186/s12909-016-0637-x PubMed PMID: 27097981; PubMed Central PMCID: PMC4839106.