Comparing Graduates Self-Assessment with Program Directors Ratings of Entrustability

Research Highlights
Brian Mavis, PhD-- Michigan State University College of Human Medicine
Chi Chang, PhD-- Michigan State University College of Human Medicine
Heather Laird-Fick, MD MPH-- Michigan State University College of Human Medicine
Angela Thompson-Busch, MD, PhD-- Michigan State University College of Human Medicine
Matthew T. Emery, MD-- Michigan State University
Dianne Wagner, MD-- Michigan State University College of Human Medicine

Purpose: A recent consensus report identified 13 foundational activities that all medical school graduates are expected to perform without direct supervision on their first day of residency.(1) This study compares graduates self-assessed confidence and program directors ratings of entrustability to perform 13 core entrustable professional activities (EPAs).

Methods: Annually, our medical school anonymously surveys graduating students about the learning environment including their confidence to perform 13 core EPAs. We also survey the program directors annually, including entrustability ratings of our graduates for the core EPAs. Data from the 2017 graduates self-assessment were compared to the 2018 ratings provided by program directors. For both groups, a rating scale ranging from 0% to 100% with 10% intervals was used.

Results: In mixed effects multivariate modelling, student bias awareness, having observed discrimination towards racial and ethnic minority patients, and ratings on tenseness of school racial climate were associated with higher interracial anxiety scores while student perception that the Eighty-nine graduating students (48%) completed the survey; 151 program directors (75%) responded. A multivariate analysis of variance was used to compare the means of two groups on each EPAs. The result was significant for the group differences (Pillais Trace = 0.537, F = 19.602, df = 13, 220, p < .001) indicating a difference in the EPA skills between graduates and program directors perspectives. The univariate F tests showed there was a significant difference between the groups for four EPAs: give or receive a patient handover to transition care responsibility (73.5 vs 79.9, p < .05); gather history and perform a physical examination (84.8 vs 78.7, p < .01); enter and discuss orders and prescriptions (52.4 vs 78.4, p < .001); perform general procedures of a physician (61.1 vs 73.3, p < .001).

Discussion: For nine EPAs, graduates confidence was comparable to program directors entrustability ratings measured a year later. For three of four EPAs where differences were found, the program directors trust exceeded the graduates confidence. The differences highlight areas for enhancing opportunities for practice and feedback in the curriculum. A limitation of this study was the comparison of unmatched cohort data; the individual responses of the students and their respective program directors could not be linked for a more robust comparison.

Significance: Assessment of entrustability is feasible, both for learners and for clinical faculty. Interestingly, students' self-assessed confidence prior to graduation is generally related to the confidence ratings of their program directors about a year later. There is a convergence of opinion from the two data sources about learner entrustability. Unlike many self-assessment tasks, students' self-assessments overall were fairly accurate, although this conclusion is based on a comparison of unmatched cohort data. These data also are useful for on-going program evaluation particularly for competency-based curricula. This assessment approach provides a strategy for gathering comparable performance data from residency programs that is meaningful to medical schools.
Authentic Assessment using EPAs: The Critical Role of Structured Professional Development for Assessors

Research Highlights
Elizabeth B. Bradley, PhD-- University of Virginia School of Medicine
James R. Martindale, PhD-- University of Virginia School of Medicine
Megan J. Bray, MD-- University of Virginia School of Medicine
Meg G. Keeley, MD-- University of Virginia School of Medicine
Christine M. Peterson, MD-- University of Virginia School of Medicine
Linda A. Waggoner-Fountain, MD, MEd-- University of Virginia School of Medicine
Maryellen E. Gusic, MD-- University of Virginia School of Medicine

Purpose: As medical schools around the world implement Entrustable Professional Activities (EPAs) as a framework for teaching and assessment of medical student performance, calls for systematic professional development have highlighted the importance of preparing supervisors for their role. (1) Determining the level of supervision a student needs requires skills in direct observation (2) and an understanding of how expectations for performance relate to standards for granting graduated autonomy. (3) Data from observation provide information for a critical dialogue in which supervisors can coach learners to facilitate continued development. (4) A comprehensive and structured approach to professional development can enable organizational change and facilitate successful implementation of a novel educational program. (5)

In this study, we describe the results of professional development programming for supervisors as one component of a comprehensive approach to prepare various stakeholders to engage in a program of EPA-based assessment and entrustment-based decision-making.

Methods: A structured, interactive professional development program to prepare faculty, residents, fellows and Master Assessors to perform EPA assessments in the clinical workplace was adapted for implementation in a variety of educational settings. Master Assessors, faculty selected to engage in assessments across disciplines and in various clinical environments, comprise the Entrustment Committee and are charged with summative decision-making about learners readiness for graduated autonomy at key transition points in the curriculum.

Content covered in the sessions includes educational theory related to competency-based assessment and EPAs. Exercises illuminate best practices in direct observation. Recordings of standardized student-patient encounters prime discussion of expectations for performance and allow participants to translate standards into decisions about supervision need: joint performance, direct or indirect supervision. Participants receive hands-on practice providing verbal feedback and documenting data from observations using web-enabled tools they will use for just in time assessment.

To assess the impact of training, differences in the supervision ratings recommended by faculty, residents/fellows, and master assessors were assessed using a series of Kruskal-Wallis (KW) non-parametric ANOVAs. The KW test (H) was used because the recommendations are ordinal and so, observed scores are replaced with rankings to test whether the sum of the rankings is different across each level of the independent variable. Separate comparisons were made for each of the EPAs measured and each learners first rating was used. All data were analyzed using SPSS, version 25.

Results: 724 assessors attended training sessions. 6754 assessments completed between February 2017 and November 2018 were used in the analysis. No statistically significant differences were found in learners first rating among the group of faculty, residents/fellows, and master assessors for EPA 1.1 (complete history) (H=.704, df=2, p=.703), 1.2 (focused history) (H=.947, df=2, p=.623), 1.3 (complete physical) (H=1.699, df=2, p=.428), 1.4 (focused physical) (H=4.817, df=2, p=.090), 2 (differential diagnosis) (H=1.319, df=2, p=.517), and 6 (oral
presentation) (H=.183, df=2, p=.913). A statistically significant difference in ratings was found for EPA 5 (written note) (H=7.640, df=2, p=.022). Analysis of pairwise comparisons for this EPA found a statistically significant difference in ratings between residents/fellows and faculty (H=17.708, df=1, p=.040).

Discussion: Structured training resulted in assessors applying consistent standards in making supervision recommendations. Purposeful professional development provides an opportunity for assessors with various levels of experience to engage in application exercises and collegial discussion that enhances norming and the creation of a shared mental model to combat subjective and idiosyncratic assessment.

Significance: Authentic assessment of learners in the workplace requires implementation of standards that are meaningful to learners, useful to supervisors, and credible for those who are responsible for summative decision-making. Inclusion of additional resources (exemplary written notes asynchronous learning tools and data visualizations) will support and strengthen efforts to norm how standards are being applied.
Implementing Clerkship OSCE Programmatic Assessment: How many data points are needed to determine EPA competency?

Research Highlights
Danielle Roussel, MD -- University of Utah School of Medicine
Katie Lappe, MD -- University of Utah School of Medicine
Tiffany Weber, MD -- University of Utah School of Medicine
Jorie Colbert-Getz, PhD, MS -- University of Utah School of Medicine

Purpose: Student assessment has traditionally been course-based with assessment scores combined to determine a final grade. In this model, monitoring and coaching for achievement of program competencies is not prioritized; students are motivated to pass a course but may not be attuned to development of skills across the curriculum. Alternatively, a programmatic assessment approach captures performance by competency across courses, enables identification of student strengths and weaknesses in multiple areas, and facilitates coaching for performance improvement. Few studies in undergraduate medical education have measured the impact of transitioning from a course-based assessment program to programmatic assessment. As assessment data accumulates for a particular competency, confidence in high stakes decisions about student achievement should increase, but this has not been validated empirically. The purpose of this study was to investigate the relationships by Entrustable Professional Activity (EPA) domain between cumulative Objective Structure Clinical Exam (OSCE) performance across clerkships at various time points and performance on an end-of-year OSCE.

Methods: In 2017, the University of Utah School of Medicine adopted a programmatic assessment model for year 3 OSCEs. OSCEs were standardized across 7 clerkships and scored items from the standardized patient checklists and encounter note rubrics were assigned to 5 EPA-aligned performance domains: EPA 1a (history taking), EPA 1b (physical examination), EPA 2 (clinical reasoning), EPA 3 (clinical testing), EPA 5 (clinical documentation). At the end of each clerkship, students completed a 2-station clerkship specific OSCE and received an overall exam score. Every 3 months during year 3 (time points 1, 2, and 3), students received an EPA score report showing their cumulative EPA domain scores relative to the entire cohort. Cumulative clerkship OSCE EPA scores at times 1 and 3 were correlated with end of year 3 OSCE scores for 110 students.

Results: Correlations between clerkship OSCE and EOY3 OSCE scores were small (< 0.30) for all EPAs at time 1. Correlations were moderate for history taking (0.43), physical examination (0.35), and documentation (0.43) and small for clinical reasoning and testing at time 3.

Discussion: As OSCE data accumulate, we can be moderately confident in our ability to make a decision about competence in history taking, physical exam and documentation. More data points or additional types of assessment are needed for decisions about clinical reasoning and ability to interpret tests.

Historically, all students have been expected to participate in an end of year 3 must pass 7-station OSCE; in the programmatic assessment model we allowed students to opt-out of the end of year OSCE if their cumulative OSCE performance at time point 2 was above 83% with no outlier low performance in any EPA domain. With EPA domain correlation data now available we will reconsider our criteria for allowing students to opt out of the final OSCE.

Use of domain specific performance data from only one type of assessment (i.e. clerkship OSCE) is a limitation; ideally programmatic assessment includes multiple assessment types to overcome different validity shortcomings of various assessment methods. Our implementation of a programmatic assessment model for clerkship OSCEs represents a first step toward an overall
Significance: As medical schools use the EPA framework for clinical assessment, it is important to understand how many data points are needed to determine competence. The results of this study suggest the number of assessment data points needed to determine competence may vary depending on the EPA.
Which Core Entrustable Professional Activities generate the most focus in student Individualized Learning Plans?

Research Highlights
Nicole M. Deiorio, MD -- Virginia Commonwealth University School of Medicine
Brieanne Dubinsky, MEd -- Virginia Commonwealth University School of Medicine
Darshita Bhatia, MD -- Inova Fairfax Hospital
Michael Ryan, MD, MEHP -- Virginia Commonwealth University School of Medicine

Purpose: The AAMC has introduced the Core EPAs for entering residency (1), and pilot institutions have endorsed the value of workplace-based assessments (WBAs), and a longitudinal coaching relationship between trainee and faculty (2) to facilitate progression toward entrustment. In 2018, we implemented a WBA program centered on the Core EPAs in conjunction with an EPA coaching program. Using self-direction, students request WBAs of their progress on one of any EPA offered in that clerkship. In the coaching program students and coaches create Individualized Learning Plans (ILPs) around progress toward independent practice of the EPAs (3, 4). We sought to understand which EPAs were the most common target of student ILPs to better support EPA-specific learner needs, and to understand how WBAs inform resultant self-monitoring and development. We hypothesized that EPAs in which students received more WBAs would feature more prominently in their ILPs.

Methods: All 214 M3 students were required to review their WBA data and to meet with their coach in the first half of their clerkship phase of training. Students pre-assessed their readiness for independent practice of each EPA as a stimulus for ILP creation. They were then required to identify strengths, areas for improvements, and learning goals for the subsequent months of training. The coach reviewed the WBA data as well. ILPs were co-created and approved by the coaches. The primary author reviewed each student's ILP content and coded any EPAs targeted in the learning goal section. We placed this in the context of the number of WBAs obtained for each EPA for any core clerkship student rotating from April-October, 2018.

Results: 200/214 (94%) students completed an ILP. Of those 200, 78% identified at least one EPA in their ILPs, with a mean of 2.1 per student (range 1-13; mean X; mode X). A small proportion (2.5%) identified all 13 EPAs in their ILPs. The most frequently identified EPAs were: EPA 2 (36%), EPA 6 (30%), EPA 1 (28%) and EPA 3 (27%). The least frequently targeted EPAs were EPA 10 (3%), EPA 13 (5%), EPA 11 (6%), and EPA 9 (6%). When only one EPA was pinpointed, it was most commonly EPA 12.

Discussion: The most frequent EPAs identified by students in their learning goals were those in which they received the most WBAs. We hypothesize two reasons for this 1) students received the most feedback on these EPAs and therefore had the most ability to understand their gaps and frame improvement. 2) students had more opportunities to perform these EPAs and they therefore perceived those EPAs as most relevant and/or attainable. We also suspect those EPAs may be the most concrete and easy-to-understand EPAs for novice learners. EPA 10 was likely less-frequently featured as students may doubt they will get the opportunity to be meaningfully involved in the care of an emergently ill patient; EPA 13 likely seems the most abstract to learners and therefore does not commonly appear in ILPs.

Significance: Students variably integrated the individual EPAs into their ILPs. EPA frequency likely depends on availability of supporting data and familiarity with the EPA. If the goal is for learners to incorporate all of the EPAs in their ILPs, future research and student/coach development should focus on making feedback around, and progress along, the less-frequently appearing EPAs seem more attainable.