

Editor's Note: The following are the Top 3 Research in Residency Education Papers selected by the JGME and the Royal College of Physicians and Surgeons of Canada for the 2018 International Conference on Residency Education meeting in Halifax, Canada. A full listing of submitted abstracts appears online (<http://www.jgme.org/page/ICREAbstracts>). Underlined author names indicate presenting author at the conference.

Winning Paper

A Grounded Theory Study of Assessment Anchors in Postgraduate Medical Education: Forthcoming Opportunities and Ongoing Tensions

Introduction: Traditional workplace-based assessment (WBA) tools have well-documented problems. A notable reconsideration of the role and design of assessment anchors is taking place in medical education. Although research has demonstrated that entrustment-based assessment tools are more reliable than traditional ones, a dearth of evidence exists for how and why these anchors actually work. This qualitative study explored the form, function, and experience of assessment anchors, aiming to develop a theoretical framework for affordances and barriers related to the adoption of entrustment-based tools.

Methods: A 2-phase, constructivist grounded theory study analyzed semistructured interviews with residents and staff from multiple specialties. Phase 1 participants ($n = 12$) had only been exposed to traditional WBA tool rating scales. Phase 2 participants ($n = 10$) had been exposed to WBA tools using entrustability anchors. Data were analyzed iteratively over the course of a 1-year study.

Results: Five themes were expressed by participants. Entrustment anchors were described as (1) concrete and defensible; (2) promoting better feedback conversations; (3) working in multiple contexts; and (4) making it possible to use the entire scale. However, entrustment anchors (5) “leave a gap” in that they do not provide information about how a trainee is doing relative to their peers and/or an expected trajectory.

Conclusions: This richer understanding of physician and resident perspectives on entrustability scales will assist WBA developers in creating more practical and acceptable tools. Understanding these perspectives is key to developing faculty development initiatives designed to introduce and improve use of these assessment tools.

N. Dudek¹, W. Gofton¹, J. Rekman¹, A. McDougall²

¹University of Ottawa, Ottawa, ON; ²CMPPA, Ottawa, ON

An Empirical Method for Determining Sentinel Milestone Ratings to Predict Residents' Attainment of Recommended Graduation Targets

Introduction: Achievement of the recommended graduation target on milestones at the end of residency is an indicator of a resident's readiness for unsupervised practice. Biannual milestone ratings allow us to estimate the likelihood that residents may (not) reach their graduation target. This study empirically derives the likelihood by retrospectively investigating milestones data over time.

Methods: Milestones data were examined from 1336 emergency medicine (EM) residents (2013–2016) in 123 programs. For each of 22 subcompetencies, a multilevel spline regression model was applied to data for residents who did not reach the target. This regression line was used to establish a cut-off milestones rating per time of evaluation. We explored various milestones rating thresholds as part of a sensitivity analysis. Odds ratios (OR) were calculated using a multilevel logistic regression model to determine the likelihood residents

DOI: <http://dx.doi.org/10.4300/JGME-D-18-00831.1>

below the thresholds would not reach the target. Negative predictive values (NPV) were also calculated to estimate the accuracy of classification.

Results: OR and NPV increased with each assessment over time. Using the spline regression cut-off at end of postgraduate year 2, ORs ranged from 2.5 to 8.9 and NPVs from 15% to 51% among 22 subcompetencies. Lowering threshold to the nearest 0.5 milestone unit, ORs ranged from 1.4 to 10.4 and NPVs from 27% to 76%.

Conclusions: ORs and NPVs calculated from national milestones ratings provide various options to help program directors identify struggling residents earlier and determine the best time to intervene. The findings of this study need to be cross-validated using different cohorts of residents.

K. Yamazaki¹, E. Holmboe¹, S. J. Hamstra²

¹Accreditation Council for Graduate Medical Education, Chicago, IL; ²University of Ottawa, Ottawa, ON

Differences in EM Trainee Caseload and Faculty Ratings Associated With Supervising Faculty Gender

Introduction: Research has shown gender-based differences in feedback provided to residents and faculty alike. Locally, we have detected similar differences in faculty evaluations. We sought to explore if this difference persisted after correction for clinical workload.

Methods: Faculty teaching evaluations were collected from 4 teaching sites at a single Canadian academic center. Resident workload was reported as the number of patients seen with a faculty member. Evaluations of 82 faculty (28 F, 54 M) were included, and analyses of variance (ANOVAs) were performed to compare reported workload and faculty ratings.

Results: From September 2013 to February 2018, 3592 faculty evaluations by trainees were recorded. Comparing the trainee workload, there was a significant main effect of gender (ANOVA $F(1,3592) = 41.0$, $P < .0001$), with lower workload under female supervising faculty. With female faculty, 14% of learners see fewer than 5 patients (versus 11% with male faculty). The largest difference is in the proportion of learners who see > 10 patients ($F = 23\%$, $M = 32\%$). Once we adjusted for caseload, there was no statistically significant difference between male and female faculty ratings by trainees (ANOVA $F(1,3) = 0.332$, $P = .80$).

Conclusions: Trainees self-reported seeing fewer patients with female supervisors. This difference raises questions about the balance between teaching and direct patient care, as the number of patients seen may impact trainee case exposure. Quantifying trainee workload may be helpful in discerning possible confounders affecting perceived teaching efficacy in the clinical environment.

L. Cook-Chaimowitz, K. Van Diepen, K. Caners, A. Pardhan, M. Welsford, T. M. Chan
McMaster University, Hamilton, ON

Editor's Note: The following are the Top 5 Resident Papers selected by the JGME and the Royal College of Physicians and Surgeons of Canada for the 2018 International Conference on Residency Education meeting in Halifax, Canada. A full listing of submitted abstracts appears online (<http://www.jgme.org/page/ICREAbstracts>). Underlined author names indicate presenting author at the conference.

Winning Paper

Does the Operative Experience of UK General Surgery Trainees Reflect the Entrustment Model of Competency Assessment?

Introduction: General surgery trainees in the United Kingdom (UK) are expected to meet requirements for procedural competency prior to completing training. Competency assessment is changing with the introduction of the entrustment concept, in the United Kingdom and North America. This study aimed to assess the operative experience of trainees and changing supervision through training as a proxy to entrustment.

Methods: Data from the Intercollegiate Surgical Curriculum Programme (ISCP) and the eLogbook databases for all UK general surgery trainees registered from August 1, 2007, who had completed training were used. Total and index procedures (IPs) were counted and variation assessed. Operative experience (IP and supervision code) by training level was assessed.

Results: We identified 311 trainees with complete data. The mean total procedures at completion of training (including assisting codes) was 2060 (SD 535). The type of IP recorded varied through training with appendectomy the most frequently undertaken IP during first year of training (mean total procedures [MTP] = 26) and emergency laparotomy during final year of training (MTP = 52). Of IPs recorded during final year, 91.2% of appendectomies (MTP = 20), 45.7% of cholecystectomies (MTP = 24), 26.1% of emergency laparotomies (MTP = 52), and 17.3% of segmental colectomies (MTP = 15) were unsupervised. The proportion of index procedures recorded as unsupervised increased through training for all IPs ($P < .05$).

Conclusions: Type and supervision of procedures performed during general surgical training in the United Kingdom changes along the lines of an entrustment model. Mapping these changes for individual trainees using existing data may provide evidence of competency.

E. Elsev¹, J. West², G. Griffiths³, D. Humes²

¹University of Nottingham, Newark; ²University of Nottingham, Nottingham; ³Ninewells Hospital, Dundee, United Kingdom

Factors Affecting the Development of Confidence During Surgical Residency Training

Introduction: Recent literature suggests the development of surgical confidence is multifactorial and affected by both trainee- and program-specific factors. This literature is based largely on surveys and questionnaires, and approaches the topic from the perspective of a “confidence crisis.” The objective of this study was to explore and better characterize the factors affecting confidence during surgical training.

Methods: This was a qualitative research study in which we conducted semistructured interviews with 7 general surgery residents to explore their experiences of confidence. Interview transcripts were coded and analyzed using inductive strategies to determine common categories, topics, and recurring themes. Each resident received a postinterview summary of their responses.

Results: Two main categories were found to affect the confidence of surgery residents: internal and external. Internal factors incorporated personal experiences (including operative experience), personal expectations, self-perception, and individual skill development. External factors involved feedback, patient outcomes,

relationships with staff, and working within a supportive environment. Interestingly, residents discussed external social factors more than case volume, technical skills, or underlying knowledge. Residents did not feel that their personal lives (eg, marital status or having children) directly affected their surgical confidence. Regardless of the factor itself, positive experiences helped build and maintain confidence by providing feelings of reassurance, encouragement, comfort, and acceptance.

Conclusions: Surgical confidence is influenced by a range of both internal and external factors. Improving our understanding of these factors can help educators improve learning experiences for residents and accelerate their progress toward being confident, independent surgeons.

M. C. Lees, B. Zheng, L. Daniels, J. S. White

University of Alberta, Edmonton, AB

Does It Work—Pilot Implementation of 3 EPAs in Busy Medical Oncology Outpatient Clinics

Introduction: Transition to Competence by Design (CBD) requires faculty assess learners using a framework of entrustable professional activities (EPAs). Adding another task to busy outpatient clinics is a concern for faculty and learners, especially in nonprocedure-based disciplines where learners are not continuously observed.

Methods: Four PGY-4 medical oncology residents completed real-time EPA assessments over a 1-month period in academic medical oncology outpatient clinics. After EPA completion, feedback forms were collected from the resident and the assessing faculty.

Results: Twenty-one resident (91%) and 23 faculty (100%) feedback forms were collected based on 23 completed EPA encounters (8–Consult, 9–Follow up, 4–Systemic Therapy). The majority of the EPAs took less than 5 minutes to complete (Faculty 87%, Residents 95%). Faculty often used 6 to 10 milestones to provide feedback to learners, the average EPA had 11 milestones. Both faculty and learners found milestones relevant and achievable. More faculty than residents agreed or strongly agreed that the EPA was a useful tool for feedback. Residents reported that if used as a checklist, the EPA did not allow the faculty to provide better feedback. Residents were less likely to find the tool useful, and found feedback was not more specific and the EPA was useful for staff who typically provide good feedback.

Conclusions: Completion of EPAs in outpatient clinics was time efficient. Faculty found EPAs useful to give specific feedback; however, residents discordantly reported use of EPAs did not improve feedback from faculty. This suggests faculty development must include feedback and coaching beyond simply completing the EPA.

A. Smrke, H. Lim

BC Cancer Agency, Vancouver, BC

Validity, Reliability, Feasibility, and Acceptability of Using the Consultation Letter Rating Scale to Assess Written Communication Competencies Among Geriatric Medicine Postgraduate Trainees

Introduction: The implementation of Competence by Design (CBD) involves using in-training assessment tools. The modified “Consultation Letter Rating Scale,” published by the Royal College of Physicians and Surgeons of Canada, evaluates written communication competencies. This multisite project evaluates the tool’s validity, reliability, feasibility, and acceptability for use in postgraduate education in geriatric medicine.

Methods: Ten geriatric medicine trainees each provided 5 consultation letters from the 2017–2018 academic year. Letters were deidentified. Six geriatricians reviewed a standardized module, and independently

completed the tool for 50 letters in a block-randomized order. They recorded the time used to complete the tool for each letter and completed a face validity survey. Inter-letter and interrater reliability was estimated using weighted and unweighted kappa. Responses on face validity were reviewed independently by 2 authors for thematic content. Participants completed a survey on the tool's usefulness.

Results: Data from 300 assessments were collected; a very small portion (4%, $N = 12$) were incomplete. There was a high agreement among raters, with an overall multiple-rater weighted kappa of 83% (95% CI 76%–89%). High level of pair-wise agreement between raters was also observed, with minimum kappa of 73% and maximum of 98%. Strong agreement across the 5 letters was observed, with a weighted kappa of 81% (95% CI 72%–88%). An average of 4.82 minutes ($SD = 3.17$) was used to complete the tool.

Conclusions: The “Consultation Letter Rating Scale” has adequate reliability and feasibility for assessing written communication competencies in postgraduate training in geriatric medicine. Analyses of acceptability and face validity are underway.

V. Xu¹, J. Hamid², M. von Maltzahn¹, T. Izukawa³, M. Norris⁴, V. Chau⁵, B. Liu⁴, C. Wong²

¹University of Toronto, Toronto, ON; ²St. Michael's Hospital, Toronto, ON; ³Baycrest Hospital, Toronto, ON; ⁴Sunnybrook Health Sciences Centre, Toronto, ON; ⁵Mount Sinai Hospital, Toronto, ON

Feedback Frequency in Competence by Design: A Quality Improvement Initiative

Introduction: Otolaryngology–head and neck surgery (OHNS) is in the initial wave of residency programs adopting Competence by Design (CBD), a new model of competency-based medical education. The University of Toronto OHNS PGY-1 residents piloted CBD during the 2016–2017 academic year, trialing several entrustable professional activities (EPAs), the task-specific assessments in CBD. The rate of completion of EPAs was monitored and targeted for a quality improvement initiative.

Methods: Residents and faculty participated in a focus group to characterize obstacles to EPA completion and to engage the stakeholders in the issue. The initial bundled intervention—rules dictating when to seek an EPA assessment and a weekly reminder a resident to the rest of the cohort—was not successful. The second intervention consisted of a leaderboard, designed on an audit-and-feedback system, sending a weekly e-mail from the program director to all PGY-1s comparing their completion rates. The number of EPAs completed weekly per resident was measured, and change was analyzed for statistical significance using control charts.

Results: The focus groups found barriers to EPA completion related to trainee attitudes, supervisor attitudes, and the measurement tool. Motivations for completion were complex, and interventions were based on intrinsic and extrinsic motivators. The leaderboard intervention demonstrated significant improvement, increasing from a baseline of 0.25 EPAs/resident-week to 2.87 EPAs/resident-week.

Conclusions: An audit-and-feedback leaderboard improved the frequency of CBD assessment completion. Resident design of the intervention fostered the necessary engagement for the initiative to succeed. Further study will have to demonstrate ongoing stability and sustainability of this process.

N. Arnstead, E. Monteiro, P. Campisi

University of Toronto, Toronto, ON