obstetric emergencies, and adult codes. There was an average of 20 patients presenting to the ED during these events. Events included a debrief, and typically lasted 60 minutes in total. Participants included individuals from various disciplines working on shift at the time of the event. Questionnaires were administered via email following the event, in which participants were asked to rank their comfort with emergency codes before and after the simulation using two 5-point Likert scales. The data from 39 questionnaires was analyzed. T-tests were used to analyze differences in self-reported comfort scores. **Results:** Questionnaire responders included nurses (41%), respiratory therapists (26%), resident physicians (10%), paramedics (3%), attending physicians (3%), students of various disciplines (10%) and other (7%). 38% of participants reported increases in comfort following simulation when compared to prior. Using the 5-point scale, the average reported score for comfort pre-simulation was 3.59 (95% CI 3.30–3.88), and the average post-simulation score was 3.97 (95% CI 3.76–4.19, p = 0.03). **Conclusion:** Our results demonstrate that weekly interprofessional in situ simulation is feasible in a high-volume ED, and significantly improves self-reported provider comfort with the identification and management of high-acuity, low-frequency events. This warrants the implementation of this simulation design to improve staff confidence and has implications for its potential role in improving team dynamics and patient safety.
Keywords: Interprofessional, Simulation

LO82

Exploring eye-tracking technology to assess competency in point-of-care ultrasound

W. Lee, MMed, J. Chenkin, MD, MEd, Sunnybrook Health Sciences Centre, Toronto, ON

Introduction: Assessment of point-of-care ultrasound (POCUS) competency has been reliant on practical, visual and written examinations performed 1-on-1 with an examiner. These tools attempt to assess competency through subjective ratings, checklists and multiple-choice questions that are labour intensive using surrogate measures. Eye-tracking has been used on a limited basis in various fields of medicine for training and assessment. This technology explores visual processing and holds great promise as a tool to monitor training progress towards the development of expertise. We hypothesize that eye-tracking may differentiate novices and experts as they progress to become competent in interpretation of POCUS images and provide an objective measure in assessment of competency. **Methods:** Medical students, residents and attending physicians working in an academic emergency department were recruited. Participants viewed a series of 16 ultrasound video clips in a POCUS protocol for Focused Assessment using Sonography in Trauma (FAST). The gaze pattern of the participants was recorded using a commercially available eye-tracking device. The primary outcome was the gaze parameters including total gaze time in the area of interest (AOI), average time to fixation on the AOI, number of fixations in the AOI and average duration of first fixation on the AOI. Secondary outcome was the accuracy on the interpretation of the FAST scan. **Results:** Four novices and eight experts completed this study. The total gaze time in the AOI (mean +/- SD) was 76.72 +/- 18.84s among experts vs 53.64 +/- 10.33s among novices (p = 0.048), average time to fixation on the AOI was 0.561 +/- 0.319s vs 1.048 +/- 0.280s (p = 0.027), number of fixations in the AOI was 158.9 +/- 29.0 vs 121.8 +/- 17.5 (p = 0.042) and average duration of first fixation was 0.444 +/- 0.119s vs 0.390 +/- 0.024s (p = 0.402). The accuracy of the answers was 79.7 +/- 14.1% vs 45.3 +/- 21.9% (p = 0.007). **Conclusion:** In this pilot study, eye tracking shows potential to differentiate between POCUS experts and novices by their gaze patterns. Gaze patterns captured by eye tracking may not necessarily translate to cognitive processing. However, it allows educators to visualise the thought processes of the learner by their gaze patterns and provide insight on how to guide them towards competency. Future studies are needed to further validate the metrics for competency in POCUS applications.
Keywords: eye tracking, medical education, ultrasound

LO83

Effect of the transition to an entrustability scale on assessor stringency and leniency on daily encounter cards in emergency medicine

S. Dewhirst, MD, W. Cheung, MD, MMed, N. Dudek, MD, T. Wood, PhD, J. Frank, MD, MA(Ed), University of Ottawa, Department of Emergency Medicine, Ottawa, ON

Introduction: Workplace based assessments (WBAs) are integral to emergency medicine residency training. However many biases undermine their validity, such as an assessor's personal inclination to rate learners leniently or stringently. Outlier assessors produce assessment data that may not reflect the learner's performance. Our emergency department introduced a new Daily Encounter Card (DEC) using entrustability scales in June 2018. Entrustability scales reflect the degree of supervision required for a given task, and are shown to improve assessment reliability and discrimination. It is unclear what effect they will have on assessor stringency/leniency – we hypothesize that they will reduce the number of outlier assessors. We propose a novel, simple method to identify outlying assessors in the setting of WBAs. We also examine the effect of transitioning from a norm-based assessment to an entrustability scale on the population of outlier assessors. **Methods:** This was a prospective pre-/post-implementation study, including all DECs completed between July 2017 and June 2019 at The Ottawa Hospital Emergency Department. For each phase, we identified outlier assessors as follows: 1. An assessor is a potential outlier if the mean of the scores they awarded was more than two standard deviations away from the mean score of all completed assessments. 2. For each assessor identified in step 1, their learners' assessment scores were compared to the overall mean of all learners. This ensures that the assessor was not simply awarding outlying scores due to working with outlier learners. **Results:** 3927 and 3860 assessments were completed by 99 and 116 assessors in the pre- and post-implementation phases respectively. We identified 9 vs 5 outlier assessors (p = 0.16) in the pre- and post-implementation phases. Of these, 6 vs 0 (p = 0.01) were stringent, while 3 vs 5 (p = 0.67) were lenient. One assessor was identified as an outlier (lenient) in both phases. **Conclusion:** Our proposed method successfully identified outlier assessors, and could be used to identify assessors who might benefit from targeted coaching and feedback on their assessments. The transition to an entrustability scale resulted in a non-significant trend towards fewer outlier assessors. Further work is needed to identify ways to mitigate the effects of rater cognitive biases.
Keywords: assessment, entrustability, rater bias