Quarterly - Medical Education - UPdate

COUNCIL OF RESIDENCY DIRECTORS IN EMERGENCY MEDICIN

UTHealth Houston | McGovern Medical School

The Association Between USMLE Step 2 Clinical Knowledge Scores and Residency Performance: A Systematic Review and Meta-Analysis

Shirkhodaie C, Avila S, Seidel H, Gibbons RD, Arora VM, Farnan JM. Acad Med. 2023 Feb 1;98(2):264-273. PMID: 36512984

Given the change of USMLE Step 1 to pass/fail, preliminary data across specialties indicates PDs may start placing more emphasis on Step 2 Clinical Knowledge (CK) when screening applicants for residency positions. The authors of this study performed a systematic review and meta-analysis of existing literature looking at the correlation between Step 2 CK scores and other measures of resident evaluation: in-training examinations (ITE) scores, board certification exams (BCEs), and other subjective measures of residency performance. Studies were eligible if they included bivariate analysis correlating Step 2 CK scores with another measure of resident performance; 68 were ultimately included and analyzed using a variety of statistical methods. The authors took it a step further and broke the data down by specialty when able. General surgery, internal medicine, and emergency medicine were the most common specialties represented. Results showed a moderate positive correlation between Step 2 CK scores and ITE scores; the correlation was markedly more positive in non-surgical specialties. There was a weak positive correlation of Step 2 CK scores with subjective measures of resident performance in surgical specialties and no correlation for non-surgical specialties. When examining correlation with specific ACGME competencies there was a weak positive correlation with professionalism and medical knowledge, there was no correlation with patient care or interpersonal and communication skills. There was a moderate positive correlation of higher Step 2 CK scores with BCE scores and passage rates among surgical and non-surgical specialties.

The authors point out that the studies analyzed were heterogenous, of low to moderate quality, and that some specialties had insufficient or no data. Also noted is that Step 2 CK is intended for licensure and was not designed to predict resident performance. This study reinforces the idea that higher Step 2 CK scores are predictive of ITE scores and BCE passage; not surprising as all are standardized tests that reflect knowledge base coupled with test-taking skills. Step 2 CK scores only weakly correlate with other characteristics that are considered necessary for a good physician. While strong ITE scores and board passage are certainly desirable attributes for residents, this study highlights the need for additional screening tools and calls into question if PDs should place further emphasis on Step 2 CK when selecting applicants for interviews.

- Amy Stubbs, MD

Vol 5.1
Winter 2023

Q-ME-UP is an educational offering brought to you via a partnership of the **McGovern Medical School at UTHealth Houston** and the **Council of Residency Directors in Emergency Medicine**. **Q-ME-UP** editors review and summarize articles related to both Graduate Medical Education and Undergraduate Medical Education

Can Simulation Measure Differences in Task-Switching Ability Between Junior and Senior Emergency Medicine Residents?

Smith D, Miller DG, Cukor J. West J Emerg Med. 2016 Mar;17(2):149-52. PMID: 26973740

It has been well documented that emergency medicine (EM) physicians deal with multiple interruptions during a standard shift. These disruptions range from having to sign an EKG while listening to a resident's presentation to managing multiple critical patients simultaneously. The ability to task-switch has been identified as a patient care competency by the Accreditation Council for Graduate Medical Education (ACGME). Mastering this skill allows physicians to respond flexibly to a changing environment. Given that simulation is often used as a method to assess milestones, this study hypothesized that task switching abilities could be measured in residents during simulation. The primary objective was to observe how residents' task switch to address a time sensitive problem while managing a critical patient. They also hypothesized that senior residents would be able to task switch more efficiently when compared to junior residents.

Investigators conducted a multicenter, prospective, observational cohort study on a convenience sample of 91 residents at three institutions; Loma Linda University Medical Center, UMass Memorial Medical Center, and The University of New Mexico Hospital. Residents in all three years of training were included in this study; 30 post graduate year PGY1s, 32 PGY2s and 29 PGY3s. All sites used high-fidelity simulation mannequins to closely mimic an emergency room environment. Residents were first presented with a diabetic patient presenting with cough, fever and vital signs consistent with sepsis. Once pneumonia was identified on chest x-ray, the patient was made hypotensive and the residents were presented with an STEMI EKG of another patient.

Data was then collected on the residents' treatment of both the septic shock and STEMI patient..

Investigators used a Cochran-Mantel-Haenszel contingency table analysis to analyze whether PGY predicted the resident's ability to properly manage sepsis. A logistic regression was used to analyze whether PGY year predicted the resident's ability to correctly interpret an EKG. Of the 91 residents included in this study, 87% properly managed the septic patient: 90.0% PGY1s, 100% PGY2s, and 96.6% PGY3s (p=0.22). Of the 87 residents who accurately managed the septic patient, 80% correctly identified a STEMI: (86.7% PGY1s, 96.9% PGY2s, and 93.1% PGY3s; p=0.35). Lastly, of the 80% who corrected managed the septic patient and identified the STEMI on EKG, 79% provided appropriate are for the STEMI patient: (73.3% PGY1s, 93.8% PGY2s, 93.8% PGY3s; p=0.07).

This study found no statistically significant difference in EM residents' ability to task switch between managing a septic patient and identifying and managing a STEMI, regardless of PGY training. Investigators considered two explanations for this data. First, residents acquired the task switching abilities needed to simultaneously manage critically ill patients during their first year of training and second, the study lacked the discrimination power needed to find a difference. If such a difference exists, it is unclear how many distractions per hour would be needed to observe this difference. Furthermore, this study was limited by its inter-rater reliability and its applicability to the emergency department setting. Residents' abilities to task switch between different patients is a level 2 milestone. Future research should attempt to measure differences in task switching ability at or above a level 3 milestone.

- Cindy Amilcar, MD (MedEd Fellow) / Benjamin Cooper, MD, MEd

Vulnerable yet Unprotected: The Hidden Curriculum of the Care of the Incarcerated Patient

Suh MI, Robinson MD. J Grad Med Educ. 2022 Dec;14(6):655-658. PMID: 36591420

The concept of hidden curriculum and the unintended learning experiences medical students and residents experience in the emergency department (ED) plays an important role in shaping their future views and practice patterns. One area that is commonly seen in the ED is the treatment of the incarcerated patient. At baseline this population is often more ill with almost half of the 2.2 million people incarcerated reporting a chronic health condition. It was also seen during the pandemic, a higher rate of infection and mortality from Covid-19 compared with the general population. What many may not know is the second most common source of healthcare for this population is the academic medical center, which translates to high likelihood of a physician encountering an incarcerated patient.

One portion of the hidden curriculum in the care of the incarcerated population is the view that these patients are less trustworthy than others presenting to our hospitals. Further emphasized is that these patients are often wrongly classified as merely being in the ED for secondary gain. Derogatory terms such as “incarceritis” are frequently used in the setting of legitimate medical conditions.

Another hidden curriculum item brought forth by the authors is loyalty by the physician to the state or carceral system. An example is given of an incarcerated adolescent who presented with a case of tenosynovitis requiring operative washout. The guard withheld parental contact information citing a theoretical danger of violence if the parents were present. In violation of standard of care, ethical principles and hospital policy, the parents were not contacted for the surgical consent.

The final hidden curriculum item is infringement of patient’s rights despite clear laws against this behavior. The lack of privacy given to patients, who often have histories and examination taken with non-clinical guards sitting in room witnessing everything, occurs much too frequently. Many patients understandably may be reluctant to provide information with such witnesses present for fear of negative consequences. Shackled patients can sometimes be denied procedures because resistance to remove the restraints by the guards.

The authors call for a shift in the approach taken to incarcerated patients. The first is to include teaching about carceral health in equity and social determinants of health discussions. Education should also include patient’s rights, especially those in custody of law enforcement.

Given the frequent encounters, ED residents and medical students have with incarcerated patients, this article serves as a great framework on where to start a change in educating our learners on how to do better when providing care to them.

-Christopher Sampson, MD

Ethnic and Racial Differences in Ratings in the Medical Student Standardized Letter of Evaluation (SLOE)

Alvarez A, Mannix A, Davenport D, Gore K, Krzyzaniak SM, Parsons M, Miller DT, Eraso D, Monteiro S, Chan TM, Gottlieb M. J Grad Med Educ. 2022 Oct;14(5):549-553. PMID: 36274773

In 2014, the Council of Residency Directors in EM (CORD) introduced the standardized letter of evaluation (SLOE) instead of more subjective letters of recommendation. CORD designed the SLOE to decrease bias in letters of recommendation. However, in many areas of medicine, underrepresented in medicine (IUM) students have long faced disparities in clinical grading and evaluation. These authors thus undertook a multi-institutional, cross-sectional study of SLOEs from three distinct academic emergency medicine departments to ascertain if the SLOE process suffers from similar disparities.

The authors define UIM as students who self-identify as "Black, Mexican-American, mainland Puerto Rican, and Native American." Using data from the 2019-2020 Electronic Residency Application Service (ERAS) cycle, a convenience sample of all Liaison Committee for Medical Education (LCME) and Commission on Osteopathic College Accreditation (COCA) applicants to the three residency programs was reviewed. International Medical School Graduates were omitted. Applicants with multiple SLOEs had each component averaged, generating an average score for each applicant for each SLOE component.

The data for each SLOE item was analyzed using an analysis of variance (ANOVA) strategy. Of 3759 applications, 2002 unique applicants (58.8% of the total EM applicant pool) and 4717 SLOEs were analyzed after inclusion/exclusion criteria were applied. Eight hundred ninety-one unique SLOEs in the data sample were identified as coming from UIM students (18.9%). The analysis demonstrated that UIM student's mean "7 Qualifications for EM" (7QEM), "Global Assessment" (GA), and "Rank List" (RL) ratings were all lower compared with non-UIM students, with an $\eta 2=0.01$ for each. While the effect size was small, it was consistently present and suggested that the SLOE process may be affected by systemic bias.

Limitations include that this cross-sectional research relies on the AAMC's definition of UIM, and there may be racial, ethnic, or otherwise unique groups not identified based on this definition. Further, while differences in SLOE results were identified, this correlation does not imply causation, and there are likely other variables that could not be controlled for in this initial analysis. Finally, no evidence exists regarding these differences' effect on the match process. More research is required to determine if there are disparities in SLOEs and the effects of these disparities on applicants' ability to match into emergency medicine.

-Joshua Goldstein, MD (PGY-3 Chief Resident) / Christopher Freeman, MD

"Faces on a screen": A qualitative study of the virtual and in-person conference experience

Gottlieb M, Sebok-Syer SS, Bawden A, Shah M, Jordan J. AEM Educ Train. 2022 Dec 20;6(6):e10827. doi: 10.1002/aet2.10827. PMID: 365620237

Virtual and hybrid conferences have gained favor in the last few years. Now, as the COVID-19 pandemic wanes and in-person gatherings are feasible, we need to understand the strengths and weakness of these different formats.

The authors conducted semi-structured interviews with attending physicians (n=26) from various subspecialties who attended virtual, hybrid or in-person continuing professional development (CPD) conferences.

Question topics included: strengths, limitations, and successful approaches of each format.

The data obtained from interview transcripts was analyzed using a modified grounded theory approach.

See table below for a summary of the benefits and limitations for each conference format

Key take-home points

When planning a conference, the format you choose should be guided by the content and audience. This paper is a useful guide to the advantages and disadvantages of all formats. From the audience perspective in person events seem best for networking and dealing with complex topics requiring focus and engagement. Virtual conferences reduce geographic barriers to participation. Further work is needed to determine the best use of hybrid conferences.

- Subarna Adhikari, MD / Aaron Danielson, MD, MAS

	Benefits	**Limitations**
In-person conferences	·Networking and sense of	·Integration into personal life
Virtual conferences	·Better flexibility and accessibility	·Highly susceptible to technology
Hybrid conferences	·More options of access	·Challenge with synchrony of
Considerations for improving virtual and hybrid conferences	·Optimization of technology and production ·Deliberate facilitation of networking and engagement ·Deliberate selection of content best suited for these environments	

Debriefing Guiding, Intermediating, Facilitating, and Teaching (GIFT): A Conceptual Framework for Simulation Educator Roles in Healthcare

Roze des Ordons AL, Eppich W, Lockyer J, Wilkie RD, Grant V, Cheng A. Simul Healthc. 2022 Oct 1;17(5):283-292. PMID: 34839303

Debriefing is a period of facilitated conversation following a simulated medical scenario that analyzes the aspects of a learner's performance to improve their future clinical practice. Debriefing and other components of simulation-based learning have become increasingly integrated into the medical education framework and encompass an area of rapidly expanding research. Existing debriefing guidelines are commonly generalized making their application to specific challenges difficult. This qualitative observational study examines the roles that debriefers adopt during post-simulation discussions with the authors consolidating their observations into a proposed framework of debriefing guidelines.

In this study a simulated case was filmed with 3 separate outcomes: engaged learners, conflicted learners, and distressed learners. Participants with a background in simulation who had completed a foundational course in debriefing were invited to participate. The authors conducted pre-simulation interviews, simulated debriefings, and post-simulation interviews. Said authors reviewed the audio recordings from the study and then subsequently coded emerging themes. Four key debriefer roles were identified: guiding, intermediating, facilitating, and teaching (GIFT), each with their own goals and strategies.

In the guiding role debriefers both developed relationships and provided a structured format as needed. They employed several strategies including introductions between participants, overview of rules and expectations, and tailoring the experience to both the needs of the learner and complexity of the case. In the intermediating role, debriefers addressed differences in perspectives to maintain a constructive environment. By intervening in conversations and reconciling differences, debriefers helped remove hierarchical or other negative dynamics from the discussion. The facilitation role involved creating reflective conversations that incorporated thoughts, emotions, and actions. Debriefers in this role invited perspective sharing, acknowledged and validated emotional experiences, and offered support. Finally, in a teaching role, debriefers discussed performance in order to create meaningful learning. By sharing perspectives, reviewing literature, and recounting personal experiences debriefers helped link simulation to real-world knowledge.

An emphasis on psychological safety was a common motif in the responsibilities of all debriefing roles. This emphasis focused on maintaining a growth mindset, inciting curiosity for knowledge, and practicing mindful body language while being an active listener. Despite prior training in simulation, a portion of debriefers had trouble transitioning between roles resulting in a tendency to act in the role in which they were most comfortable. The study found no similarities in the flow of conversation during debriefings between different participants. Each debriefer often preferred to use a specific, familiar subset of strategies. If conversational progress was unable to be achieved using these preferred strategies, debriefers would occasionally become discouraged and as a result would end conversations prematurely. Both ambiguous roles and an underprepared individual have the potential to create stress in the role and thwart the intended outcome. By utilizing the GIFT framework discussed in this study debriefers are better equipped to handle role overload. By both minimizing the ambiguity of available roles during post-simulation sessions and having additional specific tools to facilitate introspective conversation, debriefers are more effective in strategizing ways to optimize their teaching potential during simulation.

- Kevin Chambers, DO (PGY-2) / Amrita Vempati, MD

Machine Scoring of Medical Students' Written Clinical Reasoning: Initial Validity Evidence

Cianciolo AT, LaVoie N, Parker J. Acad Med. 2021 Jul 1;96(7):1026-1035. PMID: 33637657

The ongoing discussion on use of AI in medical education seems to have received a shot in the arm with the public release of ChatGPT. The focus seems to be on how learners may use this technology to cheat contemporary assessment tools. In this interesting study, the authors put AI to use as an assessment tool; they tested whether machine scoring of medical students' diagnostic justification essays could produce automated, high-quality assessments. If successful, we may be able to reintroduce clinical reasoning assessment methods that are superior to traditional knowledge assessments such as multiple-choice questions.

M3 students at Southern Illinois University School of Medicine had traditionally completed a 600-word diagnostic justification essay after a simulated patient encounter. This was intended to be an argumentative essay that explained their thought process on how they used patient data to formulate a differential, eventually arriving at a diagnosis. The essays were manually scored by faculty and (unsurprisingly) resulted in low or variable interrater reliability.

The authors took a sample of these essays, had them re-scored by trained research assistants (medical students), and then developed a machine scoring system that used the research assistant scores as the gold standard. Ultimately, the system they developed was able to achieve moderately high correlation with the gold standard (pearson correlation 0.56 to 0.82, median 0.69).

The study was designed to produce initial validity evidence, therefore used a relatively small sample of essays. The original faculty scores were not sufficiently reliable, therefore they had to re-score the essays using trained medical students. As such the "validity" of the machine score is unclear; did the machines truly, accurately measure the student's ability to justify their diagnosis? Additionally, similar AI-based assessments have not performed well with residents (compared to medical students), perhaps because residents are assessed on more complex diagnostic situations with larger areas of ambiguity. Nevertheless, the study provides an interesting glimpse into the future of resident assessment; current technology may be able to glean novel information from resident performance data. It should prompt us to continue thinking how patient outcomes data could be linked back to resident assessment.

- Nikhil Goyal, MD

Q-ME-UP Editorial Board

Samuel D. Luber, MD, MPH
McGovern Medical School at UTHealth Houston

Benjamin Cooper, MD, MEd
McGovern Medical School at UTHealth Houston

Aaron Danielson, MD, MAS
University of California - Davis School of Medicine

Christopher Freeman, MD
University of Miami Miller School of Medicine

Nikhil Goyal, MD
Henry Ford Health System

Christopher Sampson, MD
University of Missouri - Columbia School of Medicine

Amy Stubbs, MD
University of Missouri-Kansas City School of Medicine

Carmen Wolfe, MD
TriStar Skyline Medical Center

Amrita Vempati, MD
Creighton University School of Medicine - Phoenix