WestJEM

Volume XVI, Number 6, November 2015 Open Access at www.westjem.com ISSN 1936-900X

Western Journal of Emergency Medicine:

Integrating Emergency Care with Population Health

Special Issue from CDEM and CORD

Indexed in **MEDLINE**

CDEM

CORD EM

Special Issue in Educational Research and Practice

CALIFORNIA ACEP
AMERICAN COLLEGE OF EMERGENCY PHYSICIANS

ACOEP
The American College of Osteopathic Emergency Physicians

UC Irvine Health

ACADEMY OF EMERGENCY MEDICINE CAL/AAEM California Chapter Division

A Peer-Reviewed, International Professional Journal

ACOEP
The American College of Osteopathic Emergency Physicians

More Opportunities from ACOEP!

ACOEP Presents THE EDGE '15
ORAL BOARD REVIEW

August 27 - 28, 2015
Homewood Suites, Chicago, IL

ACOEP Presents THE EDGE '15
SCIENTIFIC ASSEMBLY

October 18 - 22, 2015
Loews Portofino Bay, Universal, Orlando, FL

Vist **www.acoep.org/meetings** to register today!

Western Journal *of* Emergency Medicine:

Integrating Emergency Care with Population Health

JOURNAL FOCUS

Emergency medicine is a specialty which closely reflects societal challenges and consequences of public policy decisions. The emergency department specifically deals with social injustice, health and economic disparities, violence, substance abuse, and disaster preparedness and response. This journal focuses on how emergency care affects the health of the community and population, and conversely, how these societal challenges affect the composition of the patient population who seek care in the emergency department. The development of better systems to provide emergency care, including technology solutions, is critical to enhancing population health.

Table of Contents

Policies for peer review, author instructions, conflicts of interest and human and animal subjects protections can be found online at www.westjem.com.

Western Journal *of* Emergency Medicine:
Integrating Emergency Care with Population Health
Indexed in **MEDLINE**

Table of Contents
continued

Online Manuscripts
(Full text manuscripts available open access at http://escholarship.org/uc/uciem_westjem)

Policies for peer review, author instructions, conflicts of interest and human and animal subjects protections can be found online at www.westjem.com.

Western Journal *of* Emergency Medicine:
Integrating Emergency Care with Population Health
Indexed in **MEDLINE**

Table of Contents
continued

Policies for peer review, author instructions, conflicts of interest and human and animal subjects protections can be found online at www.westjem.com.

Western Journal *of* Emergency Medicine:

Integrating Emergency Care with Population Health

Indexed in **MEDLINE**

Official Journal of the California Chapter of the American College of Emergency Physicians, the America College of Osteopathic Emergency Physicians, and the California Chapter of the American Academy of Emergency Medicine

Available in MEDLINE, PubMed, PubMed Central, CINAHL, SCOPUS, Google Scholar, eScholarship, Melvyl, DOAJ, EBSCO, EMBASE, Medscape, HINARI, and MDLinx Emergency Med. Members of OASPA.

Editorial and Publishing Office: *West*JEM/Depatment of Emergency Medicine, UC Irvine Health, 333 The City Blvd, West, Rt 128-01, Orange, CA 92868, USA
Office: 1-714-456-6389; Email: Editor@westjem.org

Western Journal *of* Emergency Medicine:

Integrating Emergency Care with Population Health

Indexed in **MEDLINE**

Official Journal of the California Chapter of the American College of Emergency Physicians, the America College of Osteopathic Emergency Physicians, and the California Chapter of the American Academy of Emergency Medicine

Available in MEDLINE, PubMed, PubMed Central, CINAHL, SCOPUS, Google Scholar, eScholarship, Melvyl, DOAJ, EBSCO, EMBASE, Medscape, HINARI, and MDLinx Emergency Med. Members of OASPA.

Editorial and Publishing Office: *West*JEM/Depatment of Emergency Medicine, UC Irvine Health, 333 The City Blvd, West, Rt 128-01, Orange, CA 92866, USA
Office: 1-714-456-6389; Email: Editor@westjem.org

Western Journal *of* Emergency Medicine:

Integrating Emergency Care with Population Health

Indexed in **MEDLINE**

This open access publication would not be possible without the generous and continual financial support of our society sponsors, department and chapter subscribers.

Professional Society Sponsors

American College of Osteopathic Emergency Physicians

California ACEP

California Chapter Division of American Academy of Emergency Medicine

Academic Department of Emergency Medicine Subscriber

Advocate Christ Medical Center
Oak Lawn, IL

Allegheny Health Network
Pittsburgh, PA

American University of Beirut
Beirut, Lebanon

Baylor College of Medicine
Houston, TX

Baystate Medical Center/ Tufts University
Springfield, MA

Boston Medical Center
Boston, MA

Brigham and Women's Hospital Department of Emergency Medicine
Boston, MA

Brown University
Providence, RI

Carl R. Darnall Army Medical Center
Fort Hood, TX

Eastern Virginia Medical School
Norfolk, VA

Emory University
Atlanta, GA

Florida Hospital Medical Center
Orlando, FL

Good Samaritan Hospital Medical Center
West Islip, NY

Henry Ford Medical Center
Detroit, MI

Highland Hospital
Oakland, CA

INTEGRIS Health
Oklahoma City, OK

Kaweah Delta Health Care District
Visalia, CA

Kennedy University Hospitals
Turnersville, NJ

Kern Medical Center
Bakersfield, CA

Lakeland HealthCare
St. Joseph, Michigan

Lehigh Valley Hospital and Health Network
Allentown, PA

Madigan Army Medical Center
Tacoma, WA

Maimonides Medical Center
Brooklyn, NY

Maricopa Medical Center
Phoenix, AZ

Massachusetts General Hospital
Boston, MA

Mayo Clinic College of Medicine
Rochester, MN

Mt. Sinai Medical Center
Miami Beach, FL

North Shore University Hospital
Manhasset, NY

Ohio State University Medical Center
Columbus, OH

Oklahoma University
Norman, OK

Presence Resurrection Medical Center
Chicago, IL

Regions Hospital/ Health Partners Institute for Education and Research
St. Paul, MN

Robert Wood Johnson Hospital
New Brunswick, NJ

Southern Illinois University
Carbondale, IL

Stanford University
Palo Alto, CA

SUNY Downstate Medical Center
New York City, NY

SUNY Upstate Medical Center
Syracuse, NY

Temple University
Philadelphia, PA

University Hospitals Case Medical Center
Cleveland, OH

University of Arizona
Tucson, AZ

University of California Davis Medical Center
Davis, CA

University of California Irvine
Orange, CA

University of California San Diego
La Jolla, CA

University of California San Francisco
San Francisco, CA

University of California San Francisco Fresno
Fresno, CA

University of Colorado & Denver Health
Denver, CO

University of Florida
Jacksonville, FL

University of Illinois at Chicago
Chicago, IL

University of Iowa
Iowa City, IA

University of Louisville
Louisville, KY

University of Maryland
College Park, MD

University of Missouri
Columbia, MO

University of Nevada
Las Vegas, NV

University of Southern California/ Keck School of Medicine
Los Angeles, CA

University of Texas
Houston, TX

University of Utah
Salt Lake City, UT

State Chapter Subscriber

Arizona Chapter Division of the American Academy of Emergency Medicine

Florida Chapter Division of the American Academy of Emergency Medicine

Great Lakes Chapter Division of the American Academy of Emergency Medicine

Tennessee Chapter Division of the American Academy of Emergency Medicine

Uniformed Services Chapter Division of the American Academy of Emergency Medicine

Virginia Chapter Division of the American Academy of Emergency Medicine

International Society Partners

Sociedad Argentina de Emergencias

Sociedad Chileno Medicina Urgencia

Thai Association for Emergency Medicine

Emergency Medicine Association of Turkey

To become a WestJEM departmental sponsor, waive article processing fee, receive print and copies for all faculty and electronic for faculty/residents, and free CME and faculty/fellow position advertisement space, please go to http://westjem.com/subscribe or contact:

Emily DeVillers
*West*JEM Staff Liaison
Phone: 1-800-884-2236
Email: sales@westjem.org

Growing a Specialty-Specific Community of Practice in Education Scholarship

Jeffrey N. Love, MD*
Douglas S. Ander, MD†

* Georgetown University School of Medicine, Department of Emergency Medicine, Washington, D.C.
† Emory University School of Medicine, Department of Emergency Medicine, Atlanta, Georgia

Section Editor: Jeffrey Love, MD
Submission history: Submitted September 11, 2015; Accepted September 11, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.28644
[West J Emerg Med. 2015;16(6):799–800.]

Emergency medicine (EM) educators have many masters. These include our hospital administrations who expect efficient patient care reflecting the priorities of safety and quality,[1,2] the accreditation council for graduate medical education which has introduced a new competency-based standard by which our learners must be educated[3,4] and last but not least, our learners that are using new educational modalities based on expanding digital platforms.[5] To be successful, educators must satisfy each of these masters against the backdrop of increasing regulations, decreasing funding[6,7] and information technology that appears to decrease our time with patients and perhaps learners in clinical practice.[8]

Success in our mission as educators is dependent upon coming together as a community of practice driven by scholarship that provides rigorous, high-impact studies that guide our educational practices. According to Lave and Wenger, a community of practice is defined as "groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly."[9] Though there are a number of accomplished EM education researchers in the United States today, the specialty remains early in its development of such a community of practice. The Council of EM Residency Directors (CORD), Clerkship Directors of EM (CDEM) and the Society of Academic EM (SAEM) have all contributed to nascent progress in this regard.

For many years, SAEM/CDEM and CORD have offered presentations and abstracts related to education scholarship at the annual CORD Academic Assembly and SAEM meetings. Over the past seven years, the American Association of Medical Colleges and CORD have collaborated in a venture to provide a year-long experience in education research called the Medical Education Research Certification (MERC) at CORD Scholars' Program.[10,11] This program is based on the experiential learning theory of Kolb and the importance of social learning that is promoted through small group projects and multifaceted exercises. To date, 186 EM educators have participated in the program. A program evaluation of MERC at CORD over the first five years demonstrates that it has succeeded in (1) improving the education research skills and knowledge of participants, (2) resulted in subsequent research projects with peers from the program, (3) promoting the growth of a community of practice that supports its members both within the area of scholarship and beyond, and (4) contributed to participants earning new leadership positions in education.[12]

The CORD Academy of Scholarship is in the process of creating opportunities that build upon the MERC at CORD program. First and foremost, there is an ongoing effort to create a collaborative education research consortium to establish benchmarks/best practices while creating and disseminating new knowledge and related applications. In addition to serving as a central source of multicenter study administration, the consortium plans to encourage collaboration. Due to the multi-institutional design of these studies, the results are more likely to generalize to similar programs/institutions. Junior faculty will be encouraged to participate with more experienced members on projects. With increasing responsibilities and experience over time the intention is to facilitate the development of growing community of education scholars. Through these efforts both MERC at CORD and the CORD Academy continue to contribute to this developing community of practice by facilitating interested faculty's ability to overcome individual obstacles to a career in scholarship including the lack of expertise, difficulty in identifying mentors, and a supportive network that promotes the professional growth of its members.[10,13-18]

Developing a successful career in education scholarship is also dependent upon a number of factors beyond the control of the individual. Institutional barriers to professional development

are much more problematic including lack of protected time for scholarly pursuits, financial support and departmental recognition of related accomplishments. Such obstacles directly result from a lack of funding available for education scholarship. Working together, CORD and the EM Foundation have created a series of education research grants aimed at increasing the value of education research. Applications for these grants became available in October of 2015.

Most recently, CDEM and CORD have collaborated to develop and introduce this inaugural edition of the WestJEM Education Supplement. We hope to build this supplement into a regular forum that explores the breadth and depth of education scholarship as it relates to EM by sharing research findings, innovations, opinions of national experts and regular updates on state-of-the-art concepts in medical education. The ability to serve as a reviewer for this supplement provides an additional opportunity to hone interested faculty's knowledge of the theory and principles that guide quality education scholarship. By providing a forum that values and cultivates education scholarship we hope to promote its growth and integration into the fabric of our specialty. The 107 manuscripts submitted for consideration are testimony to the enthusiasm and need for such a platform in our community. Ultimately, we strive to take another step in the direction of creating a community of practice centered on scholarship that will allow EM educators to grow and prosper.

Address for Correspondence: Douglas S. Ander, MD, 49 Jesse Hill Jr. Dr., Atlanta, GA 30303. Email: dander@emory.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Leape LL and Berwick DM. Five Years After To Err Is Human: What Have we Learned. *JAMA.* 2005;293:2384-90.
2. IOM Report: Crossing the Quality Chasm: A New Health System for the 21st Century. Available at: http://books.nap.edu/openbook.php?record_id=10027. Accessed Jul 13, 2015.
3. Swing SR. The ACGME Outcome Project: Retrospective and Prospective. *Med Teach.* 2007;29:648-54.
4. Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system rationale and benefits. *N Engl J Med.* 2012;366(11):1051-6.
5. Scott KR, Shy CH, Johnson NJ, et al. Integration of Social Media in Emergency Medicine Residency Curriculum. *Annals of Emerg Med.* 2014;64:396-404.
6. Martin DR, Kazzi AA, Wolford R, et al. Report from the Council of Emergency Medicine Residency Directors Subcommittee on Graduate Medical Education Funding: Effects of Decreased Medicare Support. *AEM.* 2001;8:809-14.
7. Rich EC, Liebow M, Srinivasan M, et al. Medicare Financing of Graduate Medical Education. *JGIM.* 2002;17:283-92.
8. Siegler JE, Patel NN, Dine J. Prioritizing Paperwork Over Patient Care: Why Can't We Do Both? *JGME.* 2015;7:16-8.
9. Lave J and Wenger E. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press, 1991.
10. Love JN, Coates WC, Santen SA, et al. The MERC at CORD Scholars Program in Medical Education Research: A Novel Faculty Development Opportunity for Emergency Physicians. *AEM.* 2009;16(12) Supplement 2:S37-S41.
11. Coates WC, Love JN, Farrell SE, et al. Faculty Development in Medical Education Research: A Cooperative Model between AAMC-MERC and CORD. *Acad Med.* 2010;85(5):829-36
12. Love JN, Yarris L, Santen SA , et al. A Novel, Specialty Specific Faculty Development Opportunity in Collaborative Education Research: The MERC at CORD Scholar's Program at 5 Years. *Acad Med.* In press.
13. Gruppen LD. Improving Medical Education Research. *Teach Learn Med.* 2007;19:331-5.
14. Collins J. Medical Education Research: Challenges and Opportunities. *Radiol.* 2006;240(3):639-47.
15. Zibrowski EM, Weston WW, Goldszmidt MA. "I don't have time": issues of fragmentation, prioritization, and motivation for education scholarship among medical faculty. *Med Teach* .2008;30(1):34-9.
16. Goldszmidt MA, Zibrowski EM, Weston WW. Education Scholarship: it's not just a question of "degree". *Med Teach*. 2008;30(1):34-9.
17. Yarris LM, Juve AM, Artino AR, et al. Expertise, Time, Money, Mentoring, and Reward: Systemic Barriers that Limit Education Researcher Productivity—Proceedings from the AAMC GEA Workshop. *JGME.* 2014;6(3):430-6.
18. Stenfors-Hayes T, Weurlander M, Dahlgren LO, et al. Medical Teachers' Professional Development: Perceived Barriers and Opportunities. *Teach Higher Educ.* 2010:15(4):399-408.

Emergency Medicine: On the Frontlines of Medical Education Transformation

Eric S. Holmboe, MD Accreditation Council for Graduate Medical Education, Chicago, Illinois

Section Editor: Jeffrey Love, MD
Submission history: Submitted August 11, 2015; Accepted August 11, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.28393

Emergency medicine (EM) has always been on the frontlines of healthcare in the United States. I experienced this reality first hand as a young general medical officer assigned to an emergency department (ED) in a small naval hospital in the 1980s. For decades the ED has been the only site where patients could not be legally denied care. Despite increased insurance coverage for millions of Americans as a result of the Affordable Care Act, ED directors report an increase in patient volumes in a recent survey.[1] EDs care for patients from across the socioeconomic spectrum suffering from a wide range of clinical conditions. As a result, the ED is still one of few components of the American healthcare system where social justice is enacted on a regular basis. Constant turbulence in the healthcare system, major changes in healthcare delivery, technological advances and shifting demographic trends necessitate that EM constantly adapt and evolve as a discipline in this complex environment. [West J Emerg Med. 2015;16(6):801–803.]

In this context the emergency medicine (EM) residency community has embraced the challenge to transform training for the 21st century. It is probably no accident EM was one of just five specialties to implement and report on Milestones during the 2013-2014 academic year. And it is also no surprise EM will publish the first validity study on Milestones.[2] How did EM get to this point? In this commentary, I hope to accomplish three objectives. First, I will briefly review the history of the next accreditation system (NAS), competencies and Milestones. Second, to provide an "outsider's view" of why implementation of competency-based training, and specifically the Milestones, appears to be off to a healthy start in the EM community. Finally, I will offer some thoughts on further steps necessary to realize the full potential of competency-based medical education in U.S. residency training because much work remains to be done. We are moving from a "static/stable" view of educational programs to one that is dynamic and constantly evolving. The Milestones, along with the Clinical Learning Environment Review, are regulatory representations of this shift; both are designed to be formative, continuous quality improvement components of the NAS.

A Brief History of Competencies, Milestones and the NAS

The NAS was part of the educational community's response to improve graduate medical education (GME).[3] The NAS is designed to help achieve the original vision of the Outcomes Project that was officially launched in 2001, based on the six general competencies formally approved by the Accreditation Council for GME (ACGME) and the American Board of Medical Specialties (ABMS) in 1999.[4] However, programs struggled to implement an outcomes-based approach and operationalize the competencies, new concepts to many medical educators, into meaningful changes in curriculum and assessment.

One reason for this struggle is the lack of shared mental models, or frame of reference, regarding the competencies among programs and clinical faculty. There were several reasons for this struggle. First, the ACGME/ABMS general competencies were defined in conceptual terms that were often hard to translate into practice. Second, some of the competencies, especially practice-based learning and improvement and systems-based practice, were new concepts altogether. Third, work-based assessment methods were either unavailable or not well aligned with the purpose and goals of the competencies.

The Milestones were developed collaboratively by each specialty to create the core blueprint, or roadmap, of the discipline in narrative, developmental language. In other words, they have helped to describe the competencies in more understandable language. The Milestones serve as a framework

to inform and guide curriculum, choice of assessment methods and instruments, and assessment judgments by the clinical competency committee.[5] Milestones also begin to move us away from an over-reliance on the quantification of competence, traditionally represented by numeric rating scales, toward a more qualitative, descriptive approach.

There are two important caveats. First, milestones do not define the totality of any discipline, including EM. Milestones are key elements of a larger "whole" of clinical competence. Second, substantial professional judgment, on the part of the faculty, is critical in the overall assessment of readiness for clinical practice. Informed judgment, based on multiple assessments through various forms of observation, is a cornerstone of a competency-based system.[6] Using Milestones to guide and perform systematic measurement holds promise to enhance our ability to assure the public of the effectiveness of GME to prepare physicians for practice.

Milestones and EM Training

EM was one of the early adopters in the NAS, being just one of five specialties to report both mid-year and end-year Milestones data in the 2013-14 academic year. Several aspects of the EM approach to developing and implementing the Milestones are noteworthy. First, the Milestones are grounded in the "Model of the Clinical Practice of EM" (EM Model). The EM Model consists of three core components: 1) individual conditions; 2) physician tasks; and 3) acuity levels. The knowledge, skills and attitudes comprising the current EM Model were informed by a national survey of 9,740 physicians in 2007 regarding EM practice.[7] Second, the American Board of EM (ABEM) engaged over 60% of EM residencies as part of a national validation study. Participating program directors essentially took the EM Milestones for a "test drive" and provided the EM Milestones working group with feedback on the Milestone placements and descriptions. In fact, of the final 227 Milestones included in the 23 subcompetencies, 46 Milestones were reassigned different performance levels based on the program feedback.[2,7]

The EM approach to Milestone development appears to have had an important impact. The initial validity study of the EM Milestones reported in year one shows very promising results. Factor analysis of the national Milestones data revealed a three-factor structure concordant with the three component EM Model of practice.[2] In addition, reliability coefficients for the Milestones were robust.[2] My hypothesis for these early positive findings brings me back to one of the initial purposes of the Milestones: to create shared mental models of the general competencies not only within EM residencies, but equally important between EM residencies across the country.[7] EM will likely stand as an exemplar for other specialties in how to build national standards for judging EM residents. In essence, the EM developmental approach is akin to a nationally-based performance dimension training (PDT) exercise. PDT is an established approach to helping to improve performance evaluations by getting all evaluators "on the same page."[6,8] By incorporating empiric evidence into EM certification design along with robust involvement of educators and program directors, the EM educational community has already likely made substantial progress in creating shared mental models of EM training and assessment.

Where Next?

The NAS is built on a foundational principle of continuous quality improvement. In the United States and most of the world, education and healthcare systems are experiencing significant change and disruption. We must continue to move away from a "static/stable" view of education and clinical care to one that is dynamic and constantly evolving. There are several implications for GME. First, changes in educational programs must become better integrated with the changes occurring in healthcare delivery and systems. Care of patients and populations is a dynamic, integrated process. As the frontline specialty of the healthcare system experiencing this disruption, EM is well positioned to lead and inform educational redesign. Second, work-based assessments will continue to grow in importance and prominence. One example of a useful technique in EM is end-of-shift encounter cards.[9] When used properly, encounter cards can enhance the quality of assessment and feedback. EM also leads the way in the development of chart stimulated recall (CSR), a validated method to assess clinical reasoning of actual patient care.[9] While not currently in widespread use, CSR and other performance-based methods (e.g. clinical indicators and patient experience) represent the next frontier in work-based assessment for EM training.

Third, the current Milestones are truly version 1.0; as in all continuous quality improvement processes some amount of change and revision will be needed. The processes used by the EM community to create version 1.0 will be invaluable to the larger educational community and capturing the detail behind these processes will be important. Finally, we want to ensure the Milestones do not create overly reductionistic assessments and curricula. Residency education will be most effective when the output is a whole physician who effectively integrates all competencies, however defined, into his or her practice. For example the EM community has developed entrustable professional activities (EPAs) to further help operationalize the competencies and milestones.[10,11] EPAs hold promise to help enhance curriculum and assessment, using milestones as "building blocks" for each EPA. Our collective goal is to produce physicians who can successfully enter unsupervised practice and continue their trajectory toward expertise and mastery. Competencies, Milestones and in the near future EPAs can serve as meaningful frameworks to help produce a talented, whole physician.

In conclusion, here is a request to the EM community: The ACGME, the ABEM and the dozens of talented EM faculty who volunteered their time, expertise, and wisdom to advance

the Milestones, need and welcome constructive feedback to continually improve the NAS and Milestones. Milestones are tools to facilitate and promote innovation and continuous improvement in GME in the Unites States, but they are not yet fully realized and will require changes and adjustments. We are entering a period of transformation that requires collectivism among all the key stakeholders and that can feel, like any change, uncomfortable. Only by working together through dialogue and across organizations can the full potential of outcomes-based medical education be realized. The EM educational community has clearly taken up this charge.

ACKNOWLEDGMENTS

A sincere thank you to Dr. Jeffrey Love for review of this commentary.

Address for Correspondence: Eric S. Holmboe, MD, Accreditation Council for Graduate Medical Education, 515 North State Street, Chicago, IL 60611. Email: eholmboe@acgme.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed that Dr. Eric S. Holmboe receives royalties from Mosby-Elsevier for a textbook on assessment and serves on the boards of Medbiquitous and NBME.

REFERENCES

1. American College of Emergency Physicians. ACEP Emergency Visits Up Since Implementation of ACA. Available at: http://newsroom.acep.org/ACEP-Emergency-Visits-Up-Since-Implementation-of-ACA. Accessed May 24, 2015.
2. Beeson M, Holmboe E, Korte R, et al. Initial Validity Analysis of the Emergency Medicine Milestones. *Acad Emerg Med.* 2015;in press.
3. Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system—Rationale and benefits. *N Engl J Med.* 2012;366:1051-6.
4. Batalden P, Leach D, Swing S, et al. General competencies and accreditation in graduate medical education. *Health Affairs (Millwood)*. 2002;21:103-11.
5. Holmboe ES. Realizing the promise of competency-based medical education. *Acad Med*. 2015;90:411-3.
6. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Med Teach.* 2010;32:676-82.
7. Beeson MS, Carter WA, Christopher TA, et al. The development of the emergency medicine milestones. *Acad Emerg Med.* 2013;20:724-9.
8. Woehr DJ and Huffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Org Psych.* 1994;67:189–205.
9. Sherbino J, Bandiera G, Frank JR. Assessing competence in emergency medicine trainees: an overview of effective methodologies. *CJEM.* 2008;10:365-71.
10. Ten Cate O. AM last page: what entrustable professional activities add to a competency-based curriculum. *Acad Med*. 2014;89:691.
11. Beeson MS, Warrington S, Bradford-Saffles A, et al. Entrustable professional activities: making sense of the emergency medicine milestones. *J Emerg Med*. 2014;47:441-52.

Education Scholarship and its Impact on Emergency Medicine Education

Jonathan Sherbino, MD, MEd McMaster University, Department of Emergency Medicine, Hamilton, Ontario, Canada

Section Editor: David Wald, DO
Submission history: Submitted May 4, 2015; Accepted September 9, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27355

Emergency medicine (EM) education is becoming increasingly challenging as a result of changes to North American medical education and the growing complexity of EM practice. Education scholarship (ES) provides a process to develop solutions to these challenges. ES includes both research and innovation. ES is informed by theory, principles and best practices, is peer reviewed, and is disseminated and archived for others to use. Digital technologies have improved the discovery of work that informs ES, broadened the scope and timing of peer review, and provided new platforms for the dissemination and archiving of innovations. This editorial reviews key steps in raising an education innovation to the level of scholarship. It also discusses important areas for EM education scholars to address, which include the following: the delivery of competency-based medical education programs, the impact of social media on learning, and the redesign of continuing professional development. [West J Emerg Med. 2015;16(6):804–809.]

INTRODUCTION

Since a massive reorganization more than a century ago, medical education in North America has remained largely unchanged.[1] However, increasing calls from the public for accountable care have been a lever to influence the reorganization of medical education. Changes in medical education include increasing attention to patient outcomes and patient safety,[2,3] restriction of resident duty hours,[4] increasing standards of physician certification,[5,6] and the reorganization of training programs around competency frameworks.[7,8]

Within emergency medicine (EM) the clinical teaching environment is becoming increasingly challenging with frequent interruptions, increasing patient volumes and complex patient presentations.[9,10] The cohort of learners being trained in emergency departments (EDs) is also rising through the expansion of EM training programs and the demand for off-service resident exposure to EM.[11] Finally, the competencies required of EM physicians are expanding as the science and practice of EM grows.

Within the context of this rapidly changing clinical and educational environment it is important to acknowledge that how a person learns remains constant. While education science and cognitive psychology is a growing field, these disciplines are simply gaining an increasing understanding of the established learning processes. Medical education is being forced to change to meet public, institutional, and clinical demands. The challenge for learners and teachers is to ensure that changes to education systems are appropriate, efficient and lead to best patient outcomes.

During this key period in medical education, education scholarship (ES) is emerging as an important academic pursuit to influence change. The purpose of this article is to define ES, articulate the different types of practice of education scholars and suggest areas of scholarship that will be influential for EM education.

DEFINING EDUCATION SCHOLARSHIP

Education research has an established, although narrow and specific, role in academic medicine. As with all types of research, education research uses the scientific method where observations inform testable hypotheses via a perpetually spiraling cycle. The outcomes of experiments lead to dynamic, explanatory theories. Recognizing the simplification of this description, education research is principally focused on discovery.

Nearly three decades ago, Boyer articulated a vision for academics that suggested the research of discovery was not an exclusive goal.[12] Rather, integration (i.e., synthesizing and making connections from difference fields or disciplines),

application (i.e., applying theory to practice) and teaching (i.e., determining best practices for learning) are equally valid goals for academic physicians.[13] In this framework, scholarship encompasses discovery, integration, application and teaching. The significance of this work is that it broadens the career and academic options of EM educators by validating contributions not typically defined by "research."

Building on this and other work,[14,15] the Canadian Association for Medical Education provides a definition of ES that addresses previous difficulties with discretely categorizing the above concepts. ES is "an umbrella term which can encompass both research and innovation in health professions education. Quality in [ES] is attained through work that is peer-reviewed, publicly disseminated and provides a platform that others can build on."[16] Again, this definition emphasizes that an important contribution to the mission of teaching hospitals, universities, and other academic organizations is the development of education innovations (e.g. novel trainee selection process, teaching method, assessment instrument, curriculum, faculty development initiative etc.), many of which provide solutions to the challenges currently facing EM education. Importantly, this definition incorporates the criteria suggested by Glassick[17] as necessary to adjudicate education work, especially innovations, as scholarly. (See Figure 1 for details.) The Academic Section of the Canadian Association of Emergency Physicians has endorsed the above definition of ES as a means to both promote and support education work.[11]

Scholarly innovation is partially analogous to knowledge translation. Evidence-based medicine has become one of the most influential forces in clinical medicine[18] because knowledge translation has allowed bench research to inform clinical practice.[19] In a similar fashion, scholarly innovation, influenced by education research, can broadly inform clinical education. The challenges facing EM education, as indicated above, require broad and coordinated attention. If each training program perpetually "re-invents the wheel" to address an education challenge, the delivery of EM education will be haphazard and variable. However, if the EM community promotes ES, rigorous solutions to common challenges can be disseminated, archived and built upon, allowing for progressive advancement of the field.

THE INFLUENCE OF TECHNOLOGY ON EDUCATION SCHOLARSHIP

The standards of scholarship are well established, but it is the ascendance of digital technologies that permit ES, particularly innovations, to take off. For example, all scholarship builds on previous theory, principles or best practices,[11] which requires a formalized search of the literature to ensure that scholarship is properly grounded and situated to advance the field. Traditionally, a literature search has meant the use of established databases (e.g. PubMed, Embase, ERIC, etc.) and medical subject headings (MEsH), none of which are well structured to support education searches. Moreover, a portion of valuable education literature is gray literature "which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers."[20] The sophisticated algorithms of web search engines (e.g., Google Scholar) allow the identification of education innovations in the gray literature that previously would not have been discovered by traditional literature search methods. This new process allows for more effective knowledge sharing and synthesis.

Peer review, the second key component of scholarship, has traditionally involved a small (e.g., two to five) group of content experts. As any editor or author knows, the quality of peer review can be quite variable and subjective, owing in part to the size of the peer-review sample. With the inclusion of comment and discussion features on journal websites and the prominence of social media-based discussion via blogs, microblogs (e.g., Twitter) and networking websites (e.g., Facebook), post-publication review of ES has rapidly developed. This crowd-sourced review process can promote a more transparent (provided reviewers identify themselves) and rigorous (provided the sample is sufficiently large to achieve saturation and has appropriate representation of content expertise) review.[21]

Finally, digital technologies provide new platforms for dissemination beyond traditional print journals and conference proceedings. Disseminating and archiving research findings or innovations are the final key criteria for scholarship. This step allows the field to advance through exposure to new work, provides the opportunity for peer review and informs future scholarship. Certainly, traditional journals have started to acknowledge that education innovation is important for inclusion (e.g., *WestJEM* Educational Advances submission category). However, open-access publishers (e.g. the Winnower, thewinnower.com), portals (e.g., MedEdPORTAL, mededportal.org), websites (e.g., Academic Life in EM Medical Education in Cases series, aliem.com/medic/), blogs (e.g. *iTeachEM*, iteachem.net), podcasts (e.g. *KeyLIME* - Key Literature in Medical Education https://itunes.apple.com/ca/podcast/keylime/id594247091?mt=2), and mobile apps (e.g., CanMEDS Springboards for Emergency Physicians, itunes.apple.com/ca/app/canmeds-springboards-for-emergency/id555109611?mt=8) are all examples where digital technology provides alternate platforms to print media for disseminating and archiving innovations. In many of these instances the influence of the free, open-access medical education (FOAM) movement, developed within the EM community, improves audience accessibility when compared with print (i.e., subscription-based) journals.[22]

DEVELOPING EDUCATION SCHOLARS

There are three general categories of education scholars. The original education scholars are education researchers in the traditional mold. With graduate level training, protected time and academic trajectories based on successful grant

Criteria
• Clear goals (i.e., defined and specific innovation that meets a need)
• Adequate preparation (i.e., builds on theory, principles or best practics in the area)
• Appropriate methods (i.e., proper and effective design)
• Significant results (i.e., meaningful contribution to the field)
• Effective presentation (i.e., broad dissemination)
• Reflective critique (i.e., analysis or evaluation of the innovation)

Figure 1. Criteria for assessing an education innovation as scholarship.[17]

applications and publication records, this group follows a traditional academic pathway, where content expertise is education science.[23]

The second, and largest, group is frontline EM teachers who are invested in a particular project and want to take the necessary steps to raise their work to the standards of scholarship. Unlike education research, the barrier to education innovation is low; thus, support for these "grassroots" efforts can yield significant returns. When an education innovation is raised to the standard of scholarship the teacher benefits from increased academic recognition. For the EM education community, when education innovations are appropriately developed and widely disseminated common needs are often addressed, preventing reduplication of effort, and decreasing the collective work of EM educators. Ten key steps to raise an education innovation to the standard of scholarship are detailed in Figure 2.[24]

However, to address the complex, multi-faceted practical issues currently facing EM education will require a dedicated group of education scholars who contribute in a sustained and directed, rather than opportunistic, fashion. This third group of clinician educators, in contrast to the other two groups, is committed to clinical practice to ensure familiarity with the issues facing education programs, grounded in education theory, and regularly producing scholarship (particularly innovations) that systematically addresses the complex needs of the EM education community.[25] A clinician educator occupies a unique space between education researchers and frontline teachers, formally trained but focused on frontline problems that require innovations. This (presumably smaller) group requires three key elements for success.

First, formal, but practical, training allows for grounding in education theory that informs scholarship. Akin to the research graduate training necessary for a career as an education researcher, a medical education fellowship orients a dedicated clinician educator to best practices and informing theories when designing an innovative curriculum or assessment instrument. The core concepts required of such fellowships for EM education scholars have been well articulated.[26-28]

Second, dedicated clinician educators cannot excel while working in isolation. A community of practice, defined as "the collaborative, informal networks that support professional practitioners in their efforts to develop shared understandings and engage in work-relevant knowledge building,"[29] is an important element of success. Communities of practice are collectively engaged in the creation of new ideas and innovations. Communities of practice connect education scholars across institutions, leveraging collective analysis of a problem to create an innovative solution. Participation involves more than networking or exchanging data. Junior members can be fostered via legitimate peripheral participation, progressively developing abilities that allow greater contribution over time.[30] Thus, EM teachers who approach ES as opportunistic (i.e. "one-off") can be progressively engaged to more regularly contribute to the collective issues facing the EM education community.[31]

Finally, in the same manner as successful researcher, successful clinician educators require institutional support and resources. The important step here is articulating to funding agencies and institutional leadership the educational value and academic legitimacy of theory-informed, peer-reviewed, publically disseminated innovations. In one national environmental scan of institutional support for ES, it was found that only 50% of institutions specifically included education innovation as an acceptable avenue for academic promotion.[16] Another North American study showed that promotion based on ES was less valued than clinical research.[32] If the quality of medical education research correlates with the amount of funding, than possibly the same may be true for the quality of medical education innovation.[33] Bandiera et al. articulate the key elements required to support a culture of ES, including protected time, infrastructure funding, reward models for productivity, and widely accepted metrics (especially alternative metrics to traditional markers) to measure impact and outcomes.[34]

KEY EM EDUCATION SCHOLARSHIP AREAS

Recognizing that an issue explored by an education scholar will be influenced by the specifics of personal interest and local need, there are a number of emerging topics relevant to education that deserve attention from the EM education community.

Delivery of Competency-based Medical Education (CBME) Programs

CBME organizes the delivery of EM education based on the abilities (i.e. competencies) required of graduates.[35] The Accreditation Council for Graduate Medical Education/ American Board of EM Milestones Project,[7] the Association of American Medical Colleges Core Entrustable Professional Activities for Entering Residency,[36] and the CanMEDS 2015 Framework[8] are all CBME initiatives. Presumably, they have been implemented to meet the contemporary challenges of medical education. However, a lot of work is still required to make these initiatives functional. How will work-based

1. Determine if you have the time and motivation to transform your innovation into scholarship. It will not be feasible for every idea.
2. Partner with established scholars, particularly if you have limited experience. Look for partners inside and outside EM.
3. Incorporate the criteria for scholarship into your innovation at the beginning. Adding required elements at a later stage may not be possible or may limit the quality of the scholarship.
4. Review existing theories, principles, or best practices so that your innovation advances (rather than duplicates) the field.
5. Use peer review at multiple stages to refine and improve the quality of the innovation.
6. Build capacity by including junior colleagues in your projects. They may become the next generation of EM education scholars.
7. Incorporate an evaluation process so that you can determine the outcome of your innovation.
8. Discuss your innovation and evaluation process with your institutional review board before you start.
9. Use a framework to guide the development of your innovation and focus your scholarship.
10. Improve your abilities as an education scholar by participating in a community of practice, including serving as a peer reviewer.

Figure 2. Key steps to transform an education innovation into scholarship.[24]
EM, emergency medicine

assessment be designed in a valid fashion?[37] How will personalized learner advancement, based on completion of milestones, be balanced against the service needs of pre-established rotations? Will accreditation continue as a function of process (e.g. number of rotations completed, availability of hospital resources etc.) or outcomes (e.g. quality of patient care)?[38] Program directors, educators and front-line teachers equally have an opportunity to collectively contribute to the innovations that will make CBME work.

Impact of Social Media on Learning

Technology has permitted individuals from around the world to collectively share, curate and develop medical education resources.[39] The rapid adoption of social media and FOAM is highly influential in EM education. For education scholars there are numerous issues that require attention. For example, how is quality determined in a non-hierarchical, immediate-publication environment?[40,41] How can alternative metrics be used to demonstrate the impact of scholarly innovations?[42] How can just-in-time digital education resources be used to optimize bedside clinical care?[43,44]

Redesign of Continuing Professional Development / Continuing Medical Education

A third key EM ES topic addresses the ongoing maintenance of competence of physicians in practice. With the rapid advancement of medical science, the learning curves of physicians in practice can no longer be assumed to maintain as a plateau, rather, continued growth should be anticipated.[45] Moreover, while traditional approaches to continuing professional development (CPD) have focused on the individual,[46] emerging constructs view the patient care team (i.e. microsystem) as the unit of intervention.[47] For an EM education scholar, how is ongoing learning assessed and certified in practice? What innovations promote team-based, in-situ learning? Is there a role for maintenance of certification of a team?

CONCLUSION

As medical education in North America reorganizes and EM practice becomes increasingly complex, the need for ES to provide solutions and inform best practices grows. ES, including both research and innovation, is informed by theories and principles, peer reviewed and disseminated and archived for others to use. Distinct from education researchers and frontline teachers, there is a need for clinician educators to systematically develop innovations that address the common, practical needs of the education community. Finally, important areas for EM education scholars to address include the delivery of competency-based medical education programs, the impact of social media on learning, and the redesign of continuing professional development.

Address for Correspondence: Jonathan Sherbino, MD, MEd, Hamilton General Hospital, McMaster Clinic, 237 Barton St. E, Hamilton, ON, Canada, LAL 2x2. Email: sherbino@mcmaster.ca.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Irby DM, Cooke M, O'Brien BC. Calls for reform of medical education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. *Acad Med*. 2010;85(2):220-7.
2. Kalet AL, Gillespie CC, Schwartz MD, et al. New measures to establish

the evidence base for medical education: identifying educationally sensitive patient outcomes. *Acad Med*. 2010;85(5):844-51.
3. The Safety Competencies Steering Committee. *The Safety Competencies: Enhancing Patient Safety Across the Health Professions.* Ottawa: Canadian Patient Safety Institute; 2008.
4. National Steering Committee on Resident Duty Hours. *Fatigue, Risk and Excellence: Towards a Pan-Canadian Consensus on Resident Duty Hours.* Ottawa, ON: The Royal College of Physicians and Surgeons of Canada. 2013.
5. Nasca TJ, Philibert I, Brigham T, et al. The Next GME Accreditation System - Rationale and Benefits. *N Engl J Med.* 2012;366:1051-6.
6. Hawkins RE, Lipner RS, Ham HP, et al. American Board of Medical Specialties Maintenance of Certification: theory and evidence regarding the current framework. *J Contin Educ Health Prof*. 2013;33(Suppl 1):S7-19.
7. Beeson MS, Carter WA, Christopher TA, et al. The development of the emergency medicine milestones. *Acad Emerg Med*. 2013;20(7):724-9.
8. Frank JR, Snell LS, Sherbino J, et al. *Draft CanMEDS 2015 Physician Competency Framework – Series III.* Ottawa, ON: The Royal College of Physicians and Surgeons of Canada; 2014.
9. Chisholm CD, Collison EK, Nelson DR, et al. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Acad Emerg Med.* 2000;7:1239–43.
10. Chisholm CD, Weaver CS, Whenmouth L, et al. A task analysis of emergency physician activities in academic and community settings. *Ann Emerg Med.* 2011;58:117–22.
11. Sherbino J, Van Melle E, Bandiera G. Education scholarship in emergency medicine part 1: innovating and improving teaching and learning. *CJEM*. 2014;16(Suppl 1):S1-5.
12. Boyer EL. *Scholarship reconsidered: priorities for the professoriate.* Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching; 1990.
13. Steinert Y and Snell L. Educational innovation and scholarship: from curriculum design to implementation. In: Sherbino J, Frank JR, ed. *Educational design: a CanMEDS guide for the health professions*. Ottawa, ON: Royal College of Physicians and Surgeons; 2011:81-6.
14. Shulman L. The scholarship of teaching. *Change.* 1999;31(5):11.
15. Glassick CE. Scholarship assessed: evaluation of the professoriate [Special Report (Carnegie Foundation for the Advancement of Teaching)]. San Francisco, CA: Jossey-Bass; 1997.
16. Van Melle E, Lockyer J, Curran V, et al. Toward a common understanding: supporting and promoting education scholarship for medical school faculty. *Med Educ.* 2014:48:1190–200.
17. Glassick CE. Boyer's expanded definitions of scholarship, the standards for assessing scholarship, and the elusiveness of the scholarship of teaching. *Acad Med.* 2000;75:877-80.
18. Medical Milestones. The British Journal of Medicine Website. Available at: http://www.bmj.com/content/medical-milestones. Accessed May 1, 2015.
19. Lang ES, Wyer P, Tabas JA, et al. Educational and research advances stemming from the Academic Emergency Medicine consensus conference in knowledge translation. *Acad Emerg Med*. 2010;17(8):865-9.
20. GL'99 Conference Program. *Fourth International Conference on Grey Literature: New Frontiers in Grey Literature*. GreyNet, Grey Literature Network Service. Washington, D.C.: 1999.
21. Sherbino J, Arora VM, van Melle E, et al. A Definition and Criteria for Social Media-based Scholarship in Health Professions Education: Results from the International Conference on Residency Education. *Postgrad Med J* (in press).
22. Nickson CP and Cadogan MD. Free Open Access Medical education (FOAM) for the emergency physician. *Emerg Med Australas*. 2014; 26(1):76-83.
23. Biros MH, Barsan WG, Lewis RJ, et al. Supporting emergency medicine research: developing the infrastructure. *Acad Emerg Med*. 1998;5(2):177-84.
24. Bhanji F, Cheng A, Frank JR, et al. Education scholarship in emergency medicine part 3: a "how-to" guide. *CJEM*. 2014;16(Suppl 1):S13-8.
25. Sherbino J, Frank JR, Snell L. Defining the key roles and competencies of the clinician-educator of the 21st century: a national mixed-methods study. *Acad Med*. 2014;89(5):783-9.
26. Coates WC, Lin M, Clarke S, et al. Defining a core curriculum for education scholarship fellowships in emergency medicine. *Acad Emerg Med*. 2012;19(12):1411-8.
27. Yarris LM, Coates WC, Lin M, et al. A suggested core content for education scholarship fellowships in emergency medicine. *Acad Emerg Med*. 2012;19(12):1425-33.
28. Love JN, Coates WC, Santen SA, et al. The MERC at CORD Scholars Program in medical education research: a novel faculty development opportunity for emergency physicians. *Acad Emerg Med*. 2009;16(Suppl 2):S37-41.
29. Confessore SJ. Building a Learning Organization: Communities of Practice, Self-Directed Learning, and Continuing Medical Education. *J Contin Educ Health Prof*. 1997;17:5-11.
30. Sherbino J, Snell L, Dath D, et al. A national clinician-educator program: a model of an effective community of practice. *Med Educ Online*. 2010;6(15).
31. Farrell SE, Digioia NM, Broderick KB, et al. Mentoring for clinician-educators. *Acad Emerg Med*. 2004;11(12):1346-50.
32. Atasoylu AA, Wright SM, Beasley BW. Promotion criteria for clinician-educators. *J Gen Intern Med.* 2003;18(9):711-6.
33. Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002-9.
34. Bandiera G, Leblanc C, Regehr G, et al. Education scholarship in emergency medicine part 2: supporting and developing scholars. *CJEM*. 2014;16(Suppl 1):S6-S12.
35. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-45.
36. Core Entrustable Professional Activities for Entering Residency.

Faculty and Learners' Guide. Washington, DC. Association of American Medical Colleges. 2014

37. Chan T, Sherbino J, McMAP Collaborators. The McMaster Modular Assessment Program (McMAP): A Theoretically Grounded Work-Based Assessment System for an Emergency Medicine Residency Program. *Acad Med*. 2015.
38. Asch DA, Epstein A, Nicholson S. Evaluating medical training programs by the quality of care delivered by their alumni. *JAMA*. 2007;298(9):1049-51.
39. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. *Acad Med.* 2013;88:893–901.
40. Thoma B, Chan TM, Paterson QS, et al. Emergency Medicine and Critical Care Blogs and Podcasts: Establishing an International Consensus on Quality. *Ann Emerg Med*. 2015;[Epub ahead of print].
41. Pillow MT, Hopson L, Bond M, et al. Social media guidelines and best practices: recommendations from the council of residency directors social media task force. *West J Emerg Med*. 2014;15(1):26-30.
42. Thoma B, Sanders JL, Lin M, et al. The social media index: measuring the impact of emergency medicine and critical care websites. *West J Emerg Med*. 2015;16(2):242-9
43. Purdy E, Thoma B, Bednarcyzk J, et al. The use of free online educational resources by Canadian emergency medicine residents and program directors. *CJEM. 2015;*17(2):101-6.
44. Mallin M, Schlein S, Doctor S, et al. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med*. 2014;89(4):598-601.
45. Pusic MV, Boutis K, Hatala R, et al. Learning Curves in Health Professions Education. *Acad Med*. 2015. [Epub ahead of print].
46. Forsetlund L, Bjørndal A, Rashidian A, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev.* 2009;(2):CD003030.
47. Lingard L, Espin S, Evans C, et al. The rules of the game: interprofessional collaboration on the intensive care unit team. *Crit Care*. 2004;8(6):R403-8.

Morbidity and Mortality Conference in Emergency Medicine Residencies and the Culture of Safety

Emily L. Aaronson, MD*†
Kathleen A. Wittels, MD*‡
Eric S. Nadel, MD*†‡
Jeremiah D. Schuur, MD, MHS*‡

*Brigham and Women's Hospital, Department of Emergency Medicine, Boston, Massachusetts
†Massachusetts General Hospital, Department of Emergency Medicine, Boston, Massachusetts
‡Harvard Medical School, Department of Emergency Medicine, Boston, Massachusetts

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted April 20, 2015; Revision received July 25, 2015; Accepted August 3, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.26559

Introduction: Morbidity and mortality conferences (M+M) are a traditional part of residency training and mandated by the Accreditation Counsel of Graduate Medical Education. This study's objective was to determine the goals, structure, and the prevalence of practices that foster strong safety cultures in the M+Ms of U.S. emergency medicine (EM) residency programs.

Methods: The authors conducted a national survey of U.S. EM residency program directors. The survey instrument evaluated five domains of M+M (Organization and Infrastructure; Case Finding; Case Selection; Presentation; and Follow up) based on the validated Agency for Healthcare Research & Quality Safety Culture survey.

Results: There was an 80% (151/188) response rate. The primary objectives of M+M were discussing adverse outcomes (53/151, 35%), identifying systems errors (47/151, 31%) and identifying cognitive errors (26/151, 17%). Fifty-six percent (84/151) of institutions have anonymous case submission, with 10% (15/151) maintaining complete anonymity during the presentation and 21% (31/151) maintaining partial anonymity. Forty-seven percent (71/151) of programs report a formal process to follow up on systems issues identified at M+M. Forty-four percent (67/151) of programs report regular debriefing with residents who have had their cases presented.

Conclusion: The structure and goals of M+Ms in EM residencies vary widely**.** Many programs lack features of M+M that promote a non-punitive response to error, such as anonymity. Other programs lack features that support strong safety cultures, such as following up on systems issues or reporting back to residents on improvements. Further research is warranted to determine if M+M structure is related to patient safety culture in residency programs. [West J Emerg Med. 2015;16(6):810–817.]

INTRODUCTION

Background

Following the Institute of Medicine's (IOM) 1999 report To Err is Human,[1] there has been widespread support for promoting a culture of safety within healthcare organizations.[2] A key goal of the patient safety movement has been creating non-punitive systems that encourage approaching safety systematically.[3] Routine case reporting and detailed case review is an essential part of this systematic approach, and facilitates the evaluation of clinical judgment and identification of systems errors. If conducted with attention to best practices such as non-punitive review, debriefing, and follow up on systems improvements it can support building strong safety cultures in medicine.[4]

Importance

Morbidity and mortality conferences (M+Ms) hold a long

tradition in medicine and play important roles in physician education and quality improvement (QI).[5] However, the historical culture of "blame and shame" embedded in M+M conferences is at odds with the goals of educating trainees in a culture of safety.[6,7] M+M conferences structured to teach residents to systematically analyze practice using QI methods in a non-punitive environment can help enhance emergency medicine (EM) safety culture.[8] Specifically, recommended techniques include conference formats that employ anonymous case reporting, use non-punitive approaches to case review, formal debriefing of trainees with cases, and follow up of actions taken to address systems issues.[9-11] While there was one published survey studying the formats of M+M in EM residency programs,[12] this did not focus on the methods that M+M uses to approach non-punitive review, debriefing, and follow up.

Goals of this investigation

This study aimed to determine the structure and processes of M+Ms in U.S. EM residency programs. Specifically, we were interested in determining the proportion of EM residency programs with conferences that were structured to 1) provide anonymous case submission and maintain anonymity during case presentations, 2) provide debriefing with residents after conferences are complete, and 3) follow up on systems issues identified during M+M.

METHODS

Study Design

We conducted a cross-sectional survey of all U.S. EM residency programs in the Society for Academic Emergency Medicine (SAEM) directory to determine the structure of their M+M conferences. We designed a 19-item survey instrument to assess the structure of EM M+M conferences in the context of safety culture. The survey instrument evaluated five domains of M+M conferences: 1) organization and infrastructure; 2) case finding; 3) case selection; 4) presentation; and 5) follow up. The domains and culture of safety questions were based on the previously validated Agency for Healthcare Research & Quality (AHRQ) Patient Safety Survey.[13] The survey was pilot-tested on three chief residents for clarity prior to survey distribution. The survey was administered using a web-based survey tool (Survey Monkey, Palo Alto, CA) and is attached as an appendix. If a respondent did not answer all questions, their complete answers were included. Our institutional review board approved this study.

Study Setting and Population

We surveyed every EM residency listed in the SAEM directory between September and December 2013. The survey was to be completed by the individual most responsible for overseeing M+M conferences at their institution. We emailed the survey to the program director, who was instructed to complete the survey or forward it on to the individual most responsible for M+M. Non-responders received repeat email requests and follow-up phone calls to encourage completion.

Key Outcome Measures

We defined anonymity as complete when neither residents nor attending physicians involved in the case were named in the case presentation, responsible for presenting it, or asked to comment during the case presentation. We defined anonymity as partial when neither residents nor attending physicians involved in the case were named in the case presentation or were responsible for presenting it, but could be asked to comment on it. Other outcomes including debriefing and follow up on systems improvement reflect the survey questions (See APPEND).

Data Analysis

We calculated descriptive statistics and 95% confidence intervals (95% CI). Bivariate associations were assessed with the Chi-square test. $P<0.05$ was considered significant.

RESULTS

Characteristics of study subjects

In the spring of 2013, 188 active EM training programs were listed in the SAEM directory. We received 164 unique responses, but had to exclude 13 as the survey response did not provide identifying information for the specific program. We include 151 responses from specific residency programs (80% response rate). One hundred forty-six out of the 151 (97%) respondents answered all questions. The demographics of responding programs can be found in Table 1. The majority of surveys were completed by residency physician administrators (108/151, 72%), with the remaining surveys completed by ED quality or safety leaders (17/151, 11%), ED clinical directors or operations administrators (11/151, 7%), attendings with no formal title (8/151, 5%) or the ED chairs (7/151, 5%). All respondents were attending physicians.

Structure

The basic characteristics of M+M conferences are reported in Table 2. When respondents were asked to rank, in order of importance, the objectives of M+M, the most common primary objective was to discuss adverse outcomes (53/151, 35%), followed by identify systems errors (47/151, 31%), identify cognitive errors (26/151, 17%), discuss interesting cases (15/151, 10%), teach individual professional accountability (6/151, 4%), and "other" (4/151, 3%).

Case finding and selection

Cases for M+M conferences were identified by multiple methods (Table 2), with email from providers (124/146, 85%) and hospital patient safety reporting system (115/146, 79%)

Table 1. Demographics of responding emergency medicine residency programs.

	N	% (95% CIs)
Region		
Northeast	47	31 (18-44)
Midwest	41	27 (13-41)
South	41	27 (13-41)
West	22	15 (0.08-30)
Program structure		
3 Year	103	68 (59-77)
4 Year	48	32 (19-45)
Program size (total number of residents)		
0-20	17	11 (0-26)
21-40	83	55 (42-68)
41-60	48	32 (16-48)
>60	3	2 (0-75)

used most frequently. Regular review of ED deaths and return visits were done at the majority of institutions (93/146, 64% and 81/146, 55% respectively), while regular review of death after admission was performed at 41% of institutions.

The decision regarding which cases to include in the conference was made by the attending supervising the conference at 40% (61/151) of institutions and through collaboration between the resident presenting and the attending supervising the conference at 40% (60/151) of institutions. At the remaining institutions, the resident presenting the cases (10/151, 7%), QI leadership (8/151, 5%) and chief residents (6/151, 4%) decided which cases to present. Six institutions (4%) listed other mechanisms for choosing which cases to present, such as an education fellow or QI committee.

When asked to rank in order of importance, the criteria used to determine which cases were presented, the most frequently top-rated criteria was the presence of errors, regardless of patient outcome (73/150, 49%), followed by severity of outcome (42/150, 28%), interesting nature of disease (29/150, 19%) and referral by another department (6/150, 4%).

Case presentation and anonymity

The structure of case presentation varies across residencies (Table 2). There is variation related to anonymity, both in case submission and during the conference itself (Figure 1). Ten percent (15/151) of programs maintained complete anonymity during the case presentation, 21% (31/151) of program maintained partial anonymity, and 69% (105/151) did not maintain anonymity. We were unable to detect any difference between the proportion of programs with complete and partial anonymity across programs of different size, location or the length of residency training (three years vs. four).

Follow up

Forty-seven percent (71/151) of programs have a formalized process for following up on systems issues identified at M+Ms. This was not different across programs of different size, location or the length of residency training. The changes made as a result of cases presented at M+M conferences are reported back at future M+Ms at 10% (15/151) of programs, and by email or another method at 58% (88/151) of programs. The remaining 32% (48/151) of programs do not regularly report back on changes made.

Forty-four percent (67/151) of programs report that they regularly debrief with residents who have had cases discussed. When this is done, it is most often done by a member of the residency administration (39/151, 26%) and less often by a chief resident (3/151, 2%) or someone else (25/151, 17%). The proportion of programs with a formalized process in place for following up on systems issues was not different across programs of different size, location or the length of residency training. There was also no difference between these variables and the proportion of programs that have a regular debriefing for residents who have had their cases presented.

There is a system to evaluate M+M conferences in place at most institutions, with attending physicians evaluating conferences at 61% (92/151) of institutions, and residents at 66% (100/151). Fifty-two percent (79/151) of institutions report that both resident and attending physicians formally evaluate these conferences.

The majority of respondents believe that M+M conferences are of educational value to the residents (144/150, 96%) (Figure 2). Most respondents also believe that case discussion focuses on identifying systems errors (124/151, 82%) and identifying cognitive errors (109/151, 72%). Eighty-eight percent (133/151) of respondents believe that M+M contributes to the culture of safety at their institutions.

DISCUSSION

M+M conferences, a requirement of the Accreditation Counsel of Graduate Medical Education (ACGME) Resident Review Committee, serve a key quality and safety function for departments of EM across the U.S. We surveyed U.S. EM residencies and found variability in the organization and structure of these conferences. Although best practice suggests that high quality incident analysis requires robust reporting, non-punitive review, and institutionalized follow up and debriefing,[4] we found that many EM programs have not implemented these best practices in their M+M conferences.

The concept of safety cultures, born out of error analysis in Chernobyl,[14,15] has been adapted from other high-reliability industries and widely applied in healthcare. The IOM has further reinforced the essential role that physicians play in creating a strong safety culture through voluntary reporting of error.[1]

EM residency programs do not appear to have

standardized this process. Of note, one fifth of programs do not use a hospital patient safety reporting system to identify cases, and EM programs are as likely to use email submissions as they are to use their hospital's patient safety reporting systems. The risk of this practice is that it bypasses institutional safety analysis, and may leave out certain stakeholders such as nursing or other relevant departments.

This lack of structured voluntary reporting is not surprising. EM also lacks an industry-wide standard for which incidents mandate peer review, likely contributing to the variation that we found in the criteria used to determine which cases are reviewed. While the Joint Commission and state boards have standards for mandatory reporting, such as perioperative death and wrong-side surgery, there are not similar standards or guidelines of which cases EDs should be reviewing.

We did, however, find that most programs are reviewing a common set of indicators including ED deaths and return visits. However, about a fifth are not reviewing hospital patient safety reports, over a third of programs are not routinely reviewing ED deaths, and even less are reviewing deaths during the inpatient stay. As deaths are the highest-risk cases, this raises concern that systematic review of all errors is not done at many EM programs. Inpatient mortality is also increasingly important to hospitals as Medicare expands the measurement of 30-day hospital mortality and increases the amount of reimbursement that is tied to performance on this metric. Given this, inpatient deaths soon after ED admission (e.g. 48 hours) should be included in standard EM case reviews.

Anonymity during incident reporting is one technique to encourage robust reporting and a strong safety culture.[16,17] We found that anonymity during M+M is not the norm at many EM programs. Just over half of programs provide anonymous case reporting, and only 10% structure their conferences to keep both attending physicians and residents completely anonymous during case discussion. Extrapolating from the evidence supporting anonymous reporting, one could posit that anonymous case review would further reinforce a culture of safety. Given that trainees are a vulnerable population, we suspect that they may experience public review of their role in adverse events as humiliating, shameful and ultimately punitive. Indeed, a survey of trainees by Wu et al. found that trainees who publicly accept responsibility for error undergo significant emotional stress, and that these events are associated with remorse, anger, guilt, and feelings of inadequacy.[18] Interestingly, in this survey, it was also noted that residents who publicly discussed their cases were more likely to report constructive changes in their practice.

This paradox–the perception that punitive environments can foster learning–is one that has likely prevented more widespread adoption of anonymous M+M conferences. Despite the potential educational effect of a punitive behavior, the cost associated with the emotional stress and disincentive to report has lead to an industry-wide movement towards non-punitive healthcare environments. This movement represents a paradigm shift from the historical structure of case review at M+M, which focused on holding individuals personally accountable for errors, regardless of contributing factors. This "blame and shame" approach hinged on identifying an

Table 2. Characteristics of emergency medicine morbidity and mortality conferences.

	n	% (95% CIs)
Organization and infrastructure		
Conference frequency		
Weekly	17	11 (0-26)
Bi-weekly (every other week)	10	7 (0-23)
Monthly	108	72 (64-81)
Less than once monthly	16	10 (0-25)
Conference length		
Shorter than 1 hour	3	2 (0-18)
1 hour	115	76 (68-84)
2 hours	27	18 (4-33)
Longer than 2 hours	6	4 (0-20)
Case finding		
Method for case identification		
Email from providers	124	85 (79-91)
Hospitals patient safety reporting system	115	79 (72-86)
Referred from risk management	98	67 (58-76)
Regular review of deaths in ED	93	64 (54-74)
Regular review of deaths after admission	60	41 (27-53)
Regular review of return visits	81	55 (44-66)
Anonymous case submission available		
Yes	84	56 (45-67)
No	67	44 (32-56)
Case selection		
Conference oversight		
Program director	37	25 (11-39)
Associate/assistant program director	22	15 (1-30)
Director of quality	54	36 (23-49)
Other faculty	35	24 (10-38)
Criteria used to determine which cases are presented		
Presence of errors, regardless of patient outcome	73	49 (41-57)
Severity of outcome	42	28 (21-35)
Interesting nature of disease	29	19 (13-25)
Referred by another department for presentation	6	4 (1-7)

ED, emergency department

Table 2. Continued.

	n	% (95% CIs)
Presentation		
Case presenter		
Resident involved in patient's care	60	41 (29-53)
Resident who presents entire conference (not involved in patient's care)	61	42 (30-54)
Faculty involved in patient's care	7	5 (0-21)
Faculty not involved	18	12 (0-28)
Anonymity		
Maintain complete anonymity	15	10 (0-25)
Maintain partial anonymity	31	21 (7-35)
Follow up		
Single individual responsible for follow up		
Yes	76	50 (39-61)
No	75	50 (39-61)
Is there a formalized process for following up on systems issues identified at M+M		
Yes	71	47 (35-57)
No	79	53 (42-64)
Changes are made as a result of cases presented at M+M conferences reported back to residents		
Yes	103	68 (59-77)
Not regularly	48	32 (19-45)
There is regular debriefing with residents who have had their cases discussed at M+M		
Yes	67	44 (32-59)
No	84	56 (45-67)
M+M is formally evaluated by attendings		
Yes	92	61 (51-71)
No	59	39 (27-51)
M+M is formally evaluated by residents		
Yes	100	66 (57-75)
No	51	34 (21-47)

M+M, morbidity and mortality conferences

individual responsible for an adverse event and encouraging critical comments from peers and superiors on the nature of the error. However, with only 4% of programs ranking "teaching personal accountability" as their primary objective of M+M, it seems that this cultural shift has begun. We posit that although programs' objectives are aligned with non-punitive review, they have been slower to make their M+M process anonymous because of these historical traditions. Indeed, with only 44% of hospitals surveyed in the most recent AHRQ Hospital Survey on Patient Safety Culture endorsing non-punitive response to error, this is a persistent problem nationally.[19]

There are likely several reasons that M+M has been slow to adopt a non-punitive approach to case review. The individuals in charge of M+M, generally senior physicians with administrative roles, trained in the era of "blame and shame" M+M. While they may understand the importance of changing M+M culture to support non-punitive review, they may not appreciate the stress and negative feelings that having one's case publicly presented at M+M causes for the average provider, especially for trainees. Second, we suspect that some physicians believe that anonymity reduces personal accountability and do not recognize the role that it plays in creating a safe space which encourages self-reporting of cases and maximizes learning.

Indeed, even within our authorship group there is not complete agreement on the role that anonymous case review plays in creating non-punitive cultures. While three of us think an anonymous M+M is more supportive of safety culture (EA, KW, JS), one of us (EN) thinks that naming providers in M+M can improve professional development and contribute to a culture of safety if done in a supportive environment. However, we all agree that anonymity may allow for more effective engagement of audience members, allowing them to separate the error from the individuals being discussed making these errors more teachable moments. Rather than the audience focusing on why that physician made the choices s/he did, they can instead focus on the generalizable systems and cognitive issues that apply to all providers presented with a similar patient, improving the educational benefit of the discussion. Given that the majority of institutions reported that they aim for the conferences to address systems issues, fostering anonymity should support this goal. Interestingly, to our knowledge, the effect of anonymity in medical case review on creating a non-punitive culture has not been researched.

Formal debriefing with residency leadership is important to assess the resident's reaction to the adverse event, self-assessment, and development of a performance improvement plan if needed. If M+M presentations are anonymous, such debriefing can insure that trainees consider personal accountability for adverse events, an important characteristic of a professional culture. Less than half of programs, however, have regular debriefing with residents who have had their cases discussed at M+M. This represents another opportunity to use cases for physician education.

Research has shown that even highly capable individuals are prone to failure if working within a poorly designed system.[20] Understanding this, and the multi-factorial nature of errors, it is essential that there are mechanisms in place to follow up on the systems issues identified at M+M

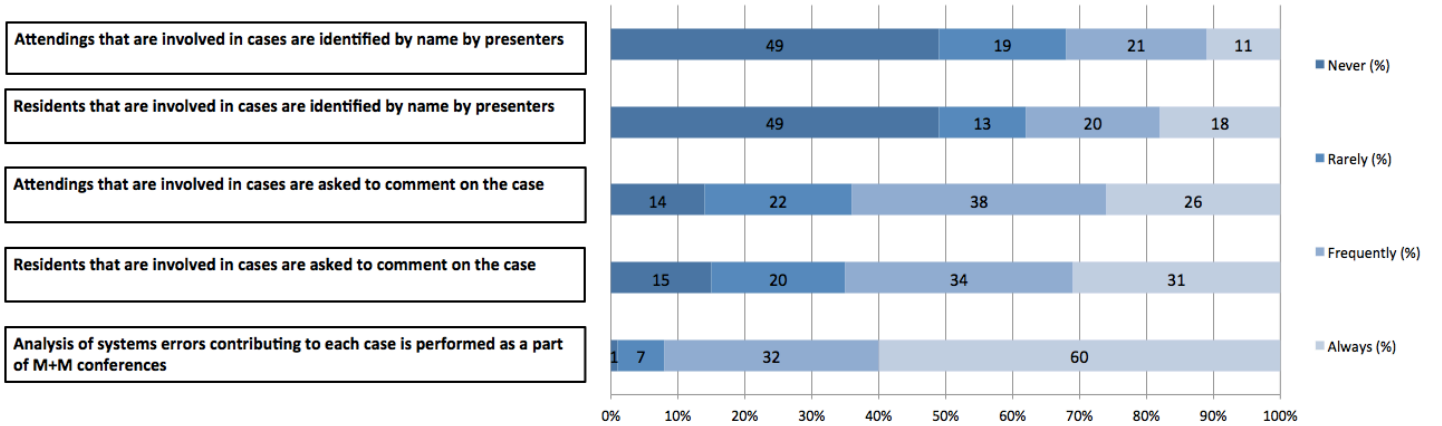


Figure 1. Structure of morbidity and mortality conferences (M+M) case presentation.

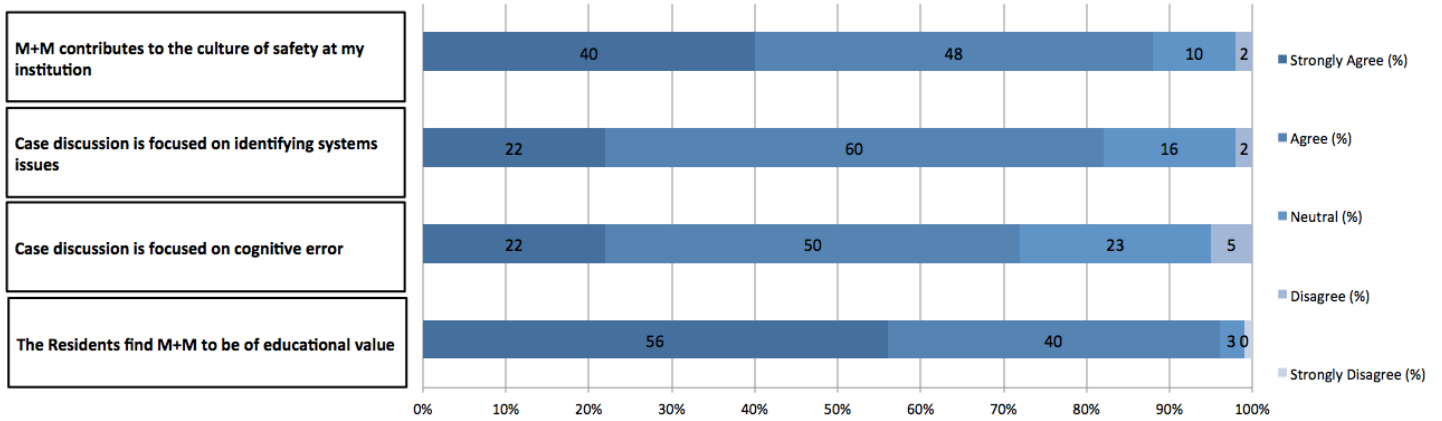


Figure 2. Goals and impact of morbidity and mortality conferences (M+M) conference.

conferences. Despite this, we found that only half of programs have a single individual who is responsible for following up on systems issues that are identified and even fewer have clear processes in place to do this. When changes are made, only a third of programs regularly report them back to staff. This may be because M+M conferences fall within the ED's educational domain rather than operations, and could be improved if M+M was integrated as a component of quality, safety and operations effort. Indeed, research has shown that by implementing a hospital-wide M+M conference with the express purpose of focusing on systems-based problems hospitals can effectively engage multiple stakeholders in the open discussion of error, identification of system failures and promotion of initiatives to improve patient safety.[21]

The lack of formalized follow up may be why only a small minority (11%) of respondents believed that systems issues identified at M+M conferences always lead to change in their EDs. It has been well documented that a leading barrier to a culture of safety is a failure to follow up with frontline providers on how adverse events led to systems improvements.[22,23] Without such follow up, frontline providers can feel that their observations about safety issues are not important, and if reported will fall on deaf ears, ultimately leading to a lower likelihood of reporting safety issues. This can deprive an ED of its most important source of information, as near misses and potential errors are much more common than adverse events that leaders usually hear about.

While we found that most EM M+Ms are not structured in a way that maximally supports strong safety cultures, we did not evaluate participants' perceptions of M+M or safety culture. Further study is needed to evaluate the impact of M+M structure on residents' perceptions of safety culture in their institutions and whether different M+M formats, such as those using anonymity, result in a less punitive culture and ultimately improved patient outcomes.

From review of this data and the literature on M+M, we recommend that programs prioritize the implementation of anonymous case reporting, creating a formalized process for follow up on the systems issues discussed and employing a clear structure for debriefing residents whose cases were discussed. Institutions can start this process by creating resident anonymity–recognizing that residents are a more vulnerable population for whom addition steps should be taken to ensure a non-punitive environment–while attendings continue to be publicly accountable for their cases. Together, these relatively basic changes will help promote robust case reporting and disclosure

of error, helping our specialty set the example for strong safety cultures in residency training and incident analysis.

LIMITATIONS

Because this was a survey, we were limited by response bias; the results represent the views of the individual respondent at those programs that responded. Our high response rate supports the external generalizability of our findings. All responses are self-reported and cannot be confirmed, another limitation of the study design. Additionally, only program directors or faculty directly responsible for M+Ms were surveyed. This results in a bias towards the opinions of those most responsible for conducting these conferences and does not represent the opinions of residents or other non-invested attending physicians. It is likely that the opinions program directors and faculty directly responsible for M+M are more positive than a general M+M audience, for example their assessment of the role of M+M on culture of safety, but also the presence of structures that would be perceived as positive, such as resident debriefing. Therefore, the proportion of residencies without structures that are perceived as positive is likely a conservative estimate. While the survey was anonymous–we did not collect individual identifiers–we did collect role and institution. Although survey respondents knew that the results would be reported anonymously, the survey was not anonymous to the study staff. This could have lead to misreporting.

As a survey, the questions asked also may not capture the nuances of many M+M conferences. For example, at some institutions cases may be presented anonymously; however, the residents and attendings involved may regularly volunteer remarks about their thought processes. These institutions would be classified as anonymous in this survey, which does not reflect the culture of safety that embraces people sharing details of error publicly, even if not asked.

CONCLUSION

This national survey of EM residencies demonstrates that while M+M conference is a standard part of EM residency training in the U.S, there is a great degree of variation in the structure of these conferences. Many programs have not integrated key tenets of a culture of safety into their M+M process, such as anonymous case review, debriefing of participants and follow up of changes that resulted from the review. While this survey could not determine the impact of M+M structure on resident education and clinical practice, it demonstrates the opportunity for EM to improve the culture of safety by incorporating these elements into regular case review in the future.

Address for Correspondence: Emily L. Aaronson, MD, Massachusetts General Hospital, Zero Emerson Place, Suite 3B, Boston, MA, 02114. Email: eaaronson@partners.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Kohn LT, Corrigan JM, Donaldson M. To Err is Human: Building a Safer Healthy System. Washington, DC: National Academy Press, Institute of Medicine; 1999.
2. Agency for Healthcare Research and Quality. AHRQ's Patient Safety Initiative: Building Foundations, Reducing Risk. An Interim Report to the United States Senate Committee on Appropriations. Rockville, MD: AHRQ; 2003.
3. Healthcare Risk Control. ECRI Institute. 2009;2:1-15.
4. McVeigh TP. Increasing Reporting of Adverse Events to Improve the Educational Value of the Morbidity and Mortality Conference. *J Am Coll Surg*. 2013;216(1):50-6.
5. Hicks CW and Makary MA. A prophet to modern medicine: Ernest Amory Codman. *BMJ*. 2013;347:7368.
6. Gordon LA. Can Cedars-Sinai's "M+M Matrix" save surgical education? *Bull Am Coll Surg.* 2004;89(6):16-20.
7. Harbison SP and Regehr G. Faculty and resident opinions regarding the role of morbidity and mortality conference. *Am J Surg.* 1999;177(2):136-9.
8. ACGME Requirements for Emergency Medicine training programs. Section IV.A.5c-f. Available at: http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/110emergencymed07012007.pdf.
9. Jones KJ, Skinner A, Xu L, et al. (2008). The AHRQ Hospital Survey on Patient Safety Culture: a tool to plan and evaluate patient safety programs. In Henriksen, K., Battles, J.B., Keyes, M.A., Grady, M.L. (Eds.), *Advances in Patient Safety: New Directions and Alternative Approaches*, Vol. 2. Culture and Redesign. AHRQ Publication No. 08-0034-2. Rockville, MD: Agency for Healthcare Research and Quality
10. Sammer CE, Lykens K, Singh KP, et al. What is Patient Safety Culture? A Review of the Literature. *J Nurs Scholarsh*. 2010;42:156–65.
11. A Review of the Safety Culture and Safety Climate Literature for the Development of the Safety Culture Inspection Toolkit. Health and Safety Executive (HSE). 2005; pp 12. Available at: http://www.hse.gov.uk/research/rrpdf/rr367.pdf.
12. Seigel TA. Morbidity and Mortality Conference in Emergency Medicine. *JEM*. 2010;38(4):507-11.
13. AHRQ Hospital Survey on Patient Safety Culture. Available at: http://www.ahrq.gov/professionals/quality-patient-safety/

patientsafetyculture/hospital/index.html.

14. Pidgeon NF. Safety Culture and Risk Management in Organizations. *J. Cross-Cultural Psychology*. 1991;22(1):129-40.
15. Glendon AL and Stanton NA. Perspectives on Safety Culture. *Safety Science*. 2000;34(1):193-214.
16. Grant MJC and Larsen GY. Effects of an anonymous reporting system on near-miss and harmful medical error reporting in a pediatric intensive care unit. *J Nurs Care Qual*. 2007;22(3):213-21.
17. Taylor JA, Brownstein D, Klein EJ, et al. Evaluation of an anonymous system to report medical errors in pediatric inpatients. *J Hosp Med.* 2007;2(4):226-33.
18. Wu AW, Folkman S, McPhee SJ, et al. Do house officers learn from their mistakes? *JAMA*. 1991;265(16):2089–94.
19. Executive Summary: 2011 User Comparative Database Report: Hospital Survey on Patient Safety Culture. April 2011. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/hospital/2011/hosp11summ.html.
20. Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. *Qual Saf Health Care*. 2006;15(3):159–64.
21. Deis JN, Smith KM, Warren MD, et al. Transforming the Morbidity and Mortality Conference into an Instrument for Systemwide Improvement. In: Henriksen K, Battles JB, Keyes MA, et al., editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 2: Culture and Redesign). Rockville (MD): Agency for Healthcare Research and Quality (US); 2008 Aug. Available at: http://www.ncbi.nlm.nih.gov/books/NBK43710/.
22. Force MV, Deering L, Hubbe J, et al. Effective strategies to increase reporting of medication errors in hospitals. *J Nurs Adm.* 2006;36(1):34-41.
23. Stump LS. Re-engineering the medication error- reporting process: removing the blame and improving the system. *Am J Health Syst Pharm.* 2000;57:S10-7.

Are Live Ultrasound Models Replaceable? Traditional versus Simulated Education Module for FAST Exam

Suzanne Bentley, MD, MPH*†
Gurpreet Mudan, MD*
Christopher Strother, MD*
Nelson Wong, MD‡

*Icahn School of Medicine at Mount Sinai, Department of Emergency Medicine, New York City, New York
†Elmhurst Hospital Center, Department of Emergency Medicine, Elmhurst, New York
‡Massachusetts General Hospital, Department of Emergency Medicine, Boston, Massachusetts

Section Editor: David Wald, DO
Submission history: Submitted April 30, 2015; Revision received August 31, 2015; Accepted September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27276

Introduction: The focused assessment with sonography for trauma (FAST) is a commonly used and life-saving tool in the initial assessment of trauma patients. The recommended emergency medicine (EM) curriculum includes ultrasound and studies show the additional utility of ultrasound training for medical students. EM clerkships vary and often do not contain formal ultrasound instruction. Time constraints for facilitating lectures and hands-on learning of ultrasound are challenging. Limitations on didactics call for development and inclusion of novel educational strategies, such as simulation. The objective of this study was to compare the test, survey, and performance of ultrasound between medical students trained on an ultrasound simulator versus those trained via traditional, hands-on patient format.

Methods: This was a prospective, blinded, controlled educational study focused on EM clerkship medical students. After all received a standardized lecture with pictorial demonstration of image acquisition, students were randomized into two groups: control group receiving traditional training method via practice on a human model and intervention group training via practice on an ultrasound simulator. Participants were tested and surveyed on indications and interpretation of FAST and training and confidence with image interpretation and acquisition before and after this educational activity. Evaluation of FAST skills was performed on a human model to emulate patient care and practical skills were scored via objective structured clinical examination (OSCE) with critical action checklist.

Results: There was no significant difference between control group (N=54) and intervention group (N=39) on pretest scores, prior ultrasound training/education, or ultrasound comfort level in general or on FAST. All students (N=93) showed significant improvement from pre- to post-test scores and significant improvement in comfort level using ultrasound in general and on FAST ($p<0.001$). There was no significant difference between groups on OSCE scores of FAST on a live model. Overall, no differences were demonstrated between groups trained on human models versus simulator.

Discussion: There was no difference between groups in knowledge based ultrasound test scores, survey of comfort levels with ultrasound, and students' abilities to perform and interpret FAST on human models.

Conclusion: These findings suggest that an ultrasound simulator is a suitable alternative method for ultrasound education. Additional uses of ultrasound simulation should be explored in the future. [West J Emerg Med. 2015;16(6):818–822.]

INTRODUCTION

Ultrasound training is an essential part of many residency programs including emergency medicine (EM), obstetrics and gynecology, surgery and internal medicine.[1-3] It is often included in the medical school clerkships within these fields. However, despite the fact that it has increasingly become a required skill for these specialties, ultrasound training varies greatly across training environments, programs, and specialties without standardized curriculums or assessment of skills[1]. For example, with the institution of the Accreditation Council of Graduate Medical Education Milestone project into EM residencies, expectations for incoming medical students have been clearly delineated.[4] A survey of EM interns reported the largest gap in training between the undergraduate curriculum and the current competency-based expectations lies in the ultrasound milestone. Only 61% of responders felt they had received the equivalent of Level 1 training in ultrasound (as opposed to 99% in professional values and team management.[5]

The expansion of the clinical indications for ultrasound highlights the potential impact of innovative ultrasound education methods. The focused assessment with sonography for trauma (FAST) is an essential scan and was chosen as the focus for this study due to its potential application across fields ,including EM, surgery and obstetrics. Traditional teaching in ultrasound is often expensive and time-consuming, requiring the use of live human models, instructors and ultrasound machines. Even with intensive resource utilization, traditional models have been shown to lag behind in the development of US interpretation skills.[6] The use of simulation for medical teaching has been shown to be feasible and useful in many different educational scenarios. A recent joint Council of Residency Directors and Academy of Emergency Ultrasound consensus document suggests simulators are a viable alternative for ultrasound training.[7] Similar to use of human simulators, the use of ultrasound simulators has been shown to be high fidelity and have the ability to enhance learning and evaluation. Simulators provide the experience of conducting an ultrasound by requiring proper probe placement and scanning techniques, and providing real-time ultrasound images as feedback. Several studies have been done validating simulators for ultrasound-guided procedures including central line placement and paracentesis.[8-10] A prior study looked at the use of a simulator in the teaching of the FAST to acquire and interpret images and showed no significant difference between the live model and ultrasound simulator groups.[11]

This study sought to evaluate the question: "are live ultrasound models replaceable?" through a more thorough education strategy and assessment requiring image acquisition and interpretation, in addition to assessment on validated American College of Emergency Physician (ACEP) ultrasound questions and clinical indications and applications of FAST use. The aim of this study was to show non-inferiority of a simulator-based ultrasound training module compared to the traditional model using human models, which is a more expensive and more time-consuming educational paradigm. For this study, it was hypothesized that medical student use of an educational module for ultrasound education using a sonographic simulator during their fourth-year EM clerkship would not be inferior to traditional teaching using live lecture followed by hands-on training with live models.

METHODS

Study Design

This was a prospective, blinded, controlled study conducted on a consecutive sample of medical students participating in a fourth-year EM clerkship. This study was approved by the institutional review board, participation was voluntary, and verbal consent was obtained from all participants.

Study Setting and Population

This study was conducted over eight months, consecutively enrolling medical students during their one month, required fourth-year medical student EM clerkship at an urban, academic, tertiary care medical center and its affiliates. No formal ultrasound education exists in the medical student curriculum at this institution. Students were randomized into a traditional training group (control group) or an ultrasound simulator group (intervention group). Randomization was based on months of the year with students rotating during odd numbered months assigned to the control group and students rotating during even numbered months assigned to the intervention group.

Study Protocol

The training was performed once a month for each group of students. Two weeks prior to the training, all students took a 20-question written pretest, composed of questions from the ACEP ultrasound question bank, a validated question bank targeting emergency department ultrasound indications, and ultrasound image interpretation. Along with the pre-test, students took a survey evaluating their baseline knowledge, prior exposure to and comfort level with ultrasound. The survey questions were based on a Likert scale of 1 through 4 with 1 representing "not at all" comfortable to 4 representing "very comfortable" with the item.

All students received a standardized, introductory lecture on the use of ultrasound, FAST basics and indications, and how to conduct a FAST exam. No student questions were answered during the lecture to maintain standardization. The lecture was delivered by the same instructor to both control and intervention groups, prior to opportunity for hands-on, self-directed practice. Noble et al demonstrated that practical training was an important part of ultrasound education.[12] The instructor demonstrated image acquisition for each of the two groups and then participants were given time for self-directed,

non-proctored practice on model of their assigned group. The control group participated in hands-on learning and practiced the FAST exam on human models (student volunteers). The intervention group participated in hands-on learning and practice of the FAST exam on the ultrasound simulator, a SonoMan Ultrasound Diagnostic Trainer (Simulab, Seattle, WA), which is a torso model with embedded electronics that simulates high fidelity normal and pathologic images in real time as the students perform ultrasound scans. No pathologic images were included or accessed during this training module.

Following the training modules, all students completed a post-test (identical to the pre-test), as well as a repeat survey. Additionally, an objective structured clinical examination (OSCE) on a live model was administered to both groups. The live model was used for the examination because the ultimate goal is to improve the ability to perform ultrasound on a live patient in a clinical setting. Students were assessed via OSCE on their performance of the FAST exam using a standardized, clinical skills "Critical Action" checklist (Figure) administered and graded by two blinded facilitators, both expert in emergency ultrasound. Examples of items on the checklist include proper probe orientation and the ability to effectively visualize each ultrasound view of the FAST exam.

Date:	Pass	Fail
1. Must wear gloves		
2. Explains criteria for FAST exam		
3. Turns on US machine and selects proper probe		
4. Obtains adequate Morison's view		
5. Obtains adequate pelvic view		
6. Obtains adequate perisplenic view		
7. Obtains adequate subcostal or parasternal long axis view		
8. Demonstrates the proper orientation of the probe marker for each view		
9. Image A. Identifies positive Morison's view.		
10. Image B. Identifies negative perisplenic view.		
11. Image C. Identifies as positive cardiac view, subcostal.		
12. Image D. Identifies as positive female pelvic view.		
13. Image E. Identifies positive perisplenic view.		
14. Image F. Identifies negative Morison's view.		
15. Image G. Identifies artifact in image as rib shadowing.		
Total:		

Figure. Critical action checklist for objective structured clinical examination of focused assessment with sonography for trauma (FAST) performance.
US, ultrasound

Outcome Measurements

Outcomes based on the following measures were evaluated: comparison of ultrasound knowledge between pre- and post-test scores in order to assess ultrasound knowledge; comparison of pre- and post-survey results of comfort with use of ultrasound; and finally the results of the OSCEs, specifically the ability of a student to perform critical actions required in order to successfully identify and interpret normal and pathologic images on FAST. We analyzed knowledge and comfort within groups from pre to post intervention, as well as between groups. OSCE scores of ability to perform ultrasounds were compared between groups.

RESULTS

All clerkship students offered participation consented and a total of 93 students were trained and tested in this study (control group N=54, intervention group N=39). There was no significant difference between groups on pre-test scores, survey results of prior ultrasound training and education or comfort level using ultrasound in general and specifically for the FAST exam. All students were in their fourth year of medical school, had similar levels of prior training in ultrasound, and similar initial comfort levels with ultrasound.

All students showed a significant improvement in their pre- and post-test scores ($p<0.001$). Mean pre-test and post-test scores for the control group were 58.5% (SD12) and 78.1% (SD 13), respectively. Mean pre-test and post-test scores for the intervention group were 56.7 % (SD13) and 75.4% (SD12). Comparison of scores between groups showed no significant difference.

Mean pre-survey comfort level was 1.38 on a four-point Likert score for the control group and 1.1 for the intervention group ($p=0.81$). Post-survey comfort level was 2.65 for the control group and 2.67 for the intervention group. All students in both control and intervention groups demonstrated significant improvement in their comfort levels using ultrasound in general and for the FAST exam after they received the intervention ($p<0.001$), with no difference between the two groups. All students reported scores of 3 or 4 on usefulness of educational session, again with no difference between groups. Additionally, there was no significant difference between groups on the OSCE standardized, clinical skills checklist conducted on the human model. Mean OSCE score was 78.2% for the control group and 81.6% for the intervention group. Overall, no difference in any of the described metrics was demonstrated between groups trained on human models versus those trained on the ultrasound simulator.

DISCUSSION

It was hypothesized that using an ultrasound simulator would not be inferior to a human model for basic ultrasound training for the FAST exam. All students showed an increase

in ultrasound knowledge, comfort and confidence after the educational intervention. In addition, the intervention group exhibited similar scores and comfort and confidence levels compared to the control group on the written, knowledge-based test and OSCE scores, which represents similar knowledge and skills gains.

Traditional ultrasound education using human models, direct faculty time, and a dedicated ultrasound machine is expensive and time consuming. Additionally, a downside to use of human models is the scarcity of pathologic examination findings. The use of an ultrasound simulator streamlines the educational process by obviating the need for human models and additional ultrasound machines for training purposes. Another notable advantage is that various pathology that would be impossible to recreate in healthy models can be demonstrated. ACEP advocates that trainees be exposed to both normal and pathologic examinations in order to increase proficiency and skill level.[13]

LIMITATIONS

Limitations to the study include a small sample size (N=93). Data was only collected at one clinical center. Outcome measures chosen demonstrate knowledge acquisition but do not offer data on clinical or patient care outcomes. Additionally, we did not assess demonstration to date of long-term retention.

CONCLUSION

The use of an ultrasound simulator is a convenient and objective method of educating medical students on ultrasound. Study results reveal that the use of a novel curriculum incorporating ultrasound simulation was non-inferior to traditional methods of ultrasound education using human models as demonstrated through knowledge-based written testing, surveys of comfort levels with ultrasound, and objective examinations of students' abilities to perform ultrasound on a human model in real time. There is a paucity of literature on the subject of validated teaching and evaluation of bedside ultrasound. This study is a proposed step on the path to developing an ultrasound curriculum using simulation methods that are non-inferior to traditional methods for teaching ultrasound. The ultimate goal is to develop an exportable and easy-to-use module for self-directed ultrasound training that will eliminate the need for models, live instructors, and that may be used across many different specialties, levels of training, and practice settings. Ultrasound simulation provides a viable solution to the problem of deliberate practice and mastery of the FAST and other ultrasound applications.

Additional uses of ultrasound simulation should be explored in the future, in particular, perhaps pioneering a validated and standardized learner-directed module for ultrasound training.

Address for Correspondence: Suzanne Bentley, MD, 7901 Broadway, Elmhurst, NY 11373. Email: bentleys@nychhc.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Ahern M, Mallin MP, Weitzel S, et al. Variability in Ultrasound Education among Emergency Medicine Residencies. *West J Emerg Med*. 2010;11(4):314-8.
2. Knudson M and Sisley A. Training Residents Using Simulation Technology: Experience with Ultrasound for Trauma. *J of Trauma* 2000;48(4):659-65.
3. Maul H, Scharf A, Baier P, et al. Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound Obstet Gynecol*. 2004;24(5):581-5.
4. Korte RC, Beeson MS, Russ CM, et al. The emergency medicine milestones: a validation study. *Acad Emerg Med*. 2013;20(7):730-5.
5. Santen SA, Rademacher N, Heron SL, et al. How competent are emergency medicine interns for level 1 milestones: who is responsible? *Acad Emerg Med*. 2013;20(7):736-9.
6. Sisley AC, Johnson SB, Erickson W, et al. Use of an Objective Structured Clinical Examination (OSCE) for the assessment of physician performance in the ultrasound evaluation of trauma. *J Trauma*. 1999;47(4):627-31.
7. Lewiss RE, Pearl M, Nomura JT, et al. CORD-AEUS: Consensus Document for the Emergency Ultrasound Milestone Project. *Acad Emerg Med* 2013;20:740-5.
8. Mendiratta-Lala M, Williams T, de Quadros N, et al. The use of a simulation center to improve resident proficiency in performing ultrasound-guided procedures. *Acad Radiol*. 2010;17(4):535-40.
9. Lee AC, Thompson C, Frank J, et al. Effectiveness of a novel training program for emergency medicine residents in ultrasound-guided insertion of central venous catheters. *CJEM*. 2009;11(4):343-8.
10. Barsuk JH, Cohen ER, Vozenilek JA, et al. J. Simulation-Based Education with Mastery Learning Improves Paracentesis Skills. *Grad Med Ed*. 2012;4(1):23-7.
11. Damewood S, Jeanmonod D, Cadigan B. Comparison of a Multimedia simulator to a Human Model for Teaching FAST Exam Image Interpretation and Image Acquisition. *Acad Emerg Med*. 2011;18:413-419.

12. Noble VE, Nelson BP, Sutingco AN, et al. Assessment of knowledge retention and the value of proctored ultrasound exams after the introduction of an emergency ultrasound curriculum. *BMC Med Educ*. 2007;7:40.
13. American College of Emergency Physicians. ACEP emergency ultrasound guidelines-2001. *Ann Emerg Med*. 2001;38:470-81.

Teaching and Assessing ED Handoffs: A Qualitative Study Exploring Resident, Attending, and Nurse Perceptions

Moira Flanigan, BA
James A. Heilman, MD
Tom Johnson, MD, PhD
Lalena M. Yarris, MD, MCR

Oregon Health and Science University, Department of Emergency Medicine, Portland, Oregon

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted April 30, 2015; Revision received August 4, 2015; Accepted August 8, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.27278

Introduction: The Accreditation Council for Graduate Medical Education requires that residency programs ensure resident competency in performing safe, effective handoffs. Understanding resident, attending, and nurse perceptions of the key elements of a safe and effective emergency department (ED) handoff is a crucial step to developing feasible, acceptable educational interventions to teach and assess this fundamental competency. The aim of our study was to identify the essential themes of ED-based handoffs and to explore the key cultural and interprofessional themes that may be barriers to developing and implementing successful ED-based educational handoff interventions.

Methods: Using a grounded theory approach and constructivist/interpretivist research paradigm, we analyzed data from three primary and one confirmatory focus groups (FGs) at an urban, academic ED. FG protocols were developed using open-ended questions that sought to understand what participants felt were the crucial elements of ED handoffs. ED residents, attendings, a physician assistant, and nurses participated in the FGs. FGs were observed, hand-transcribed, audio-recorded and subsequently transcribed. We analyzed data using an iterative process of theme and subtheme identification. Saturation was reached during the third FG, and the fourth confirmatory group reinforced the identified themes. Two team members analyzed the transcripts separately and identified the same major themes.

Results: ED providers identified that crucial elements of ED handoff include the following: 1) Culture (provider buy-in, openness to change, shared expectations of sign-out goals); 2) Time (brevity, interruptions, waiting); 3) Environment (physical location, ED factors); 4) Process (standardization, information order, tools).

Conclusion: Key participants in the ED handoff process perceive that the crucial elements of intershift handoffs involve the themes of culture, time, environment, and process. Attention to these themes may improve the feasibility and acceptance of educational interventions that aim to teach and assess handoff competency. [West J Emerg Med. 2015;16(6):823–829.]

INTRODUCTION

In order to address medical errors made during transitions of care, multiple regulatory agencies, including the Accreditation Council for Graduate Medical Education and the Joint Commission on Accreditation of Healthcare Organizations, have called for guidelines to ensure resident competency in performing safe, effective handoffs.[1,2] Successful implementation of a standardized educational intervention to improve handoffs has recently been shown to reduce medical errors in the inpatient pediatric setting.[3] In the

emergency department (ED), several studies and consensus statements have begun to identify the core elements of ED provider handoffs.[4-11] Although understanding what makes ED handoffs unique is crucial from an operational perspective, this core content knowledge does not alone shed light upon the best way to teach and assess handoff practices in the ED setting.

Educational interventions must go beyond presenting core content in order to translate knowledge into performance improvement–learners must also be receptive to the intervention and incorporate the content and process into their practice. A first step of designing instructional materials is to describe the problem that the intervention aims to address, and assess the perceptions and needs of the learner, as well as the important stakeholders who will interact with the learner in the clinical environment.[12] Although the literature describes key elements of ED handoffs, we identified a knowledge gap in understanding how the perceptions of providers' needs in the ED handoff process may impact the success of interventions designed to teach and assess ED handoffs.

To explore provider perceptions of ED handoffs that may inform or impact educational interventions aiming to improve provider performance, we designed a qualitative study to explore two central questions: 1) What themes emerge when exploring nurse, resident, attending, and midlevel provider perceptions regarding ED handoffs? and 2) What interprofessional and cultural processes take place during the ED handoff process? Using this analytical background, our study aimed to identify the crucial elements of ED-based handoffs that may be barriers to developing and implementing successful educational interventions to teach and assess ED handoff competency.

METHODS

Settings and Participants

This qualitative study of provider perceptions of ED handoffs was conducted in an urban, tertiary-care, academic ED with approximately 50,000 patient visits per year. Our three-year emergency medicine residency program, comprised of 33 residents and 24 core faculty, provides 24-hour resident coverage in the ED. Residents, attendings, charge nurses, and occasionally midlevel providers contribute to a resident-led handoff at each change of shift. All residents, attendings, midlevel providers, and nurses in the ED were invited to participate in the study by email invitation. Two of the four focus groups (FGs) were composed of a mixed group of residents, attendings and charge nurses. One of these FGs included a physician assistant. The other two FGs included only residents and attendings. FG size ranged from four to eight individuals and each group had participants who had not previously participated. Participation was voluntary and confidential. This study was approved by our institutional review board.

Study Protocol

This study was a pre-planned separate phase of a larger study that aimed to adapt a standardized handoff process to the ED setting. We used a grounded theory approach and constructivist/interpretivist paradigm that sought to understand the perceptions of the various care providers in the ED.[13-16] Our approach applied an iterative process, theoretical sampling, and a constant comparative method of data analysis. The primary phenomenon that we aimed to explore–the current intershift handoff–was studied by eliciting interprofessional perceptions regarding its standardization, safety, efficiency, and factors that may impact efforts to teach and assess handoff competency in ED providers. Our study team members included ED attendings, residents, and a student volunteer. Because multiple study investigators were known to the participants, and already a part of the handoff culture, a member of the team (MF) who was not known to the participants and was new to the culture was trained and led the facilitation of all FGs. We chose a theoretical sampling strategy that was purposive in that we sought to recruit groups of interprofessional providers representative of the providers who are actively engaged in handoffs in our institution. Subsequent theoretical sampling was guided by the categories and concepts that emerged in initial data collection, in order to maximize our understanding of relationships between concepts and developing themes. For example, the first FG did not include a midlevel provider, and early data analysis suggested that the midlevel provider voice may lend crucial insight into the handoff phenomenon. Therefore, a midlevel provider was recruited for a subsequent FG.

FG Protocol

The FG protocol for this study phase was created simultaneously with the portion of the protocol that aimed to inform standardization ED handoff practices. Open-ended questions were developed that sought to understand what participants felt were the crucial elements of ED handoffs. The discussions were allowed to proceed organically, and the facilitator probed as necessary to explore factors relevant to understanding the barriers and promoters of effective ED handoffs. Participants were prompted to rely on their cumulative experiences in all the EDs in which they had worked so that themes would have increased external validity and not be institution-specific. Three primary FGs were conducted in October 2014; each was observed, audio-recorded, and hand-transcribed. The study team conducted ongoing data analysis to determine that saturation was reached after the third FG. A fourth confirmatory FG was held, which did not reveal additional themes.

Data Analysis

Data analysis began with the hand transcription of FG proceedings during the FGs on poster paper, and then subsequent transcriptions of the audio-recordings by the MF. We anonymized and de-identified participant data. Data were

then separately analyzed and coded by two team members using an iterative process of code categorization, concept identification, and constant comparative theme and subtheme identification. In order to assure trustworthiness and credibility of data analysis, member checking was performed at the conclusion of each FG by directed group review of the data scribed onto poster paper during the session. We performed triangulation by comparing FG transcripts with observer notes and hand-transcribed session notes.

RESULTS

Analysis of FG data demonstrated three major categories that contribute to the collective conceptual understanding of the ED provider handoff: 1) the ontological framework; 2) cultural expectations; and 3) environmental factors specific to the ED setting (Figure). FG participants' perceptions revealed four dominant themes: Culture (the ability of a new educational process to change existing cultural expectations and norms, or the overall efficacy of implementation in the face of those cultural norms), Time (as seen in a collective desire for shortened, yet effective, processes and the general reticence for processes that may elongate the formal handoff procedure); Environment (how the physical location of the sign-out affects participants' learning experience, as well as the physician-patient relationship); and Process (information flow and order, consensus building). The Table illustrates the themes, subthemes, representative quotes, and educational considerations that emerged from analyzing our participants' perceptions.

Culture

An underlying culture marked by individuality, attending-resident hierarchy, and unyielding norms was alluded to throughout all FGs. Cultural reticence towards standardization in part centered on perceived lost individuality, related to both personal preference for handoff style as well as individual learning style and ability. Individuality also contributed to a complex attending-resident dynamic, in which instruction is not always based on standardized format or learner

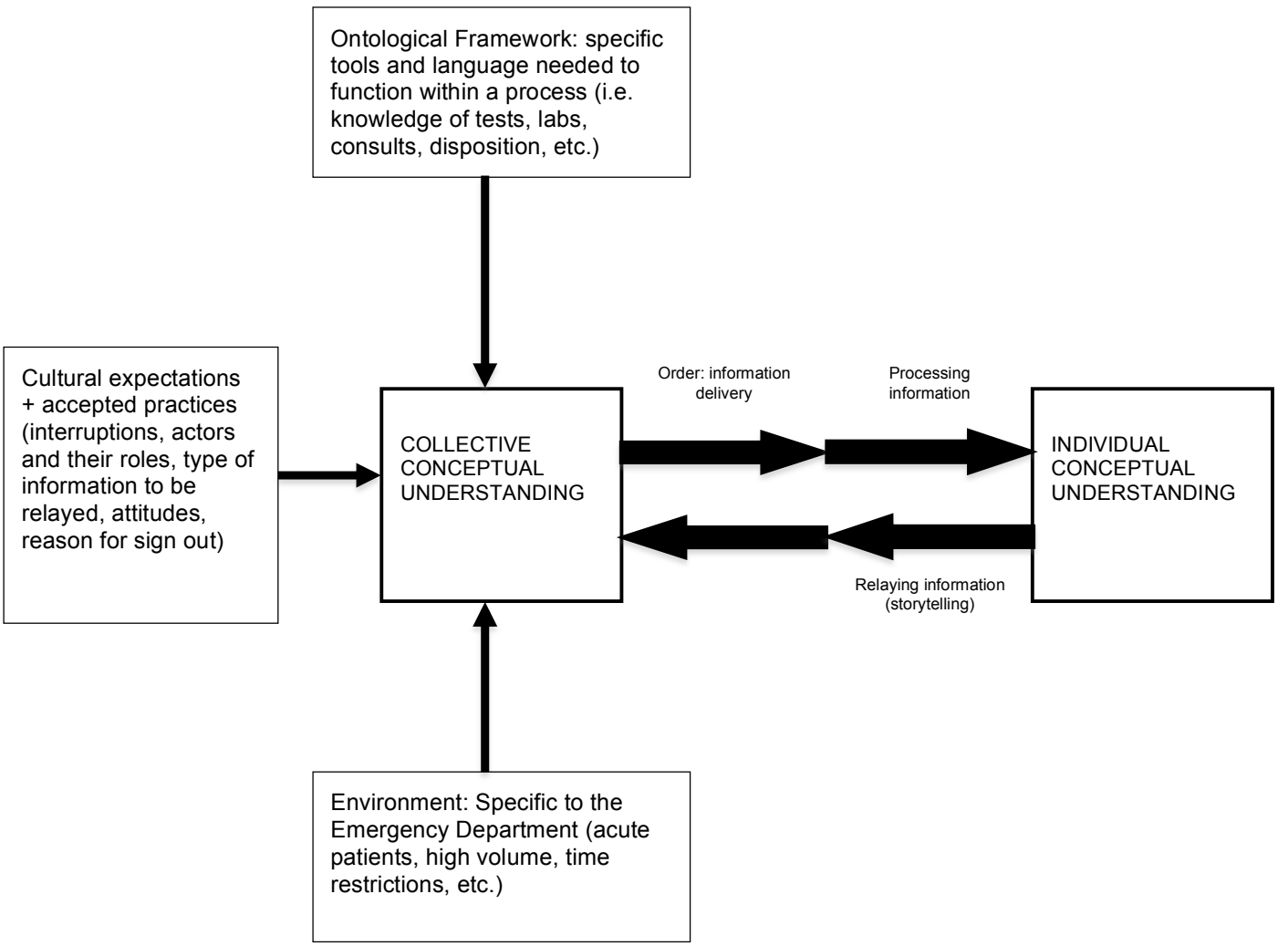


Figure. Individual and collective understanding in consensus building processes.

Table. Themes, subthemes, representative quotes, and educational considerations with regard to teaching and assessing handoffs in the emergency department.

Themes	Representative quotes	Educational considerations
Culture		
Individuality and educational interventions	"This is why it's important for it not be standardized. Because when your standardized sign-out is given to [one attending], and that's not the way he likes it, then you have to follow your [standardized way] and that's not he how he likes it."[17]	Importance of developing faculty support for the standardized handoff process
Perceptions of cultural power dynamics and educational interventions	"It's not standard. Some attendings take over all the tasks, some residents take over all the tasks, it just depends on how they decide to do it… It can be vague. You can leave the sign out process and not be sure who is going to do what."[19] "There have been some [attendings] who are more like the, 'you're here to take sign out from me, so you come find me, not the other way around'… And I hope that's an attitude that's not too prevalent."[18]	Teach and practice closed loop communication between resident and attending. Set clear departmental guidelines for resident and attending responsibilities of tasks after handoff.
Perceptions of new systems, learning, and educational interventions	"It's very dependent on who's giving the sign-out. Because we get more efficient as we mature through this career. So, an intern may be a little too verbose and add detail that's not important, or they may kind of forget a few of the important things."[18] "It, again, is… how good you are at signing out and really honing in on pertinent things."[18] "I think in general we've all learned a certain way in medicine that we like to hear things based on, obviously this is a more succinct presentation, but it just helps to tell a story, and I think during sign-out a lot of people throw that all out the window and are all over the place."[17] "It's like, it's like a recipe for making cookies, right? You can have a great recipe and still eff it up."[18]	Education on how to present a handoff in the ED setting and clinical simulation or case based practice sessions. Importance of having a standardized way to present a handoff in order to teach resident learners. Feedback from senior residents and attendings on how to improve handoff presentation.
Changing systems and improved outcomes	"But I really strongly feel that if we changed the expectations…what is appropriate to be done surrounding sign-out, then we can use our department effectively, and we can do it anywhere in the department if there is clear communication about what's going on."[17] "Less misses. It's, I mean, the spirit of a sign-out really is mostly a safety issue, not an efficiency issue, though it would be ideal to be efficient at the same time. But I think that a sign-out is for safety."[18]	Education on time management of ED patient load and preparing for handoffs. Set clear departmental guidelines for what tasks should not be signed out.
Location		
Provider perception of specific sign-out location	"I think there's a lot of interruptions because of the location."[17] "It's loud. It's loud for the patients. It's loud for the doctors. And it lends itself to interruptions."[19]	Designated sign-out locations perceived as necessary to effective hand-off process.
Provider perception of sign-out location and interprofessional relationships	"I'm not sure if any of you were here when the doctors used to be in [a dedicated room], that was where they sort of lived, and it was not good in that the nurses weren't free to go in there and ask questions and they were separated.. I think [a dedicated room] is dangerous…. Because there are people who will not come out of there."[19] "Well for me, because I'm new, it's nice to be able to ask that attending face to face."[18] "It doesn't make sense for all of the other nurses come around and the department come to a grinding halt when you're scheduling patients all over the department."[17]	Recognition of sign-out location affecting both how providers interact/learn from each, as well as how they learn about and interact with patients. Efficacy of educational intervention tied to space.
Provider perception of sign-out location and provider-patient relationships	"I think patients and families would be way more satisfied if we handed-off in the room. Because how many times do we say 'Okay Mr. Smith, oh you got tummy pain? We're going to do all these things here, and we'll come let you know when they're done'. And then you go home, and then some stranger comes in, and they're like, 'Who are you?'. Like, 'Oh, I'm Dr. So-and-So, taking over for So-and-So, and I heard so-and-so, and this is that, and the other thing.'"[18] "Where I trained we did bedside report, and… I think it cut down on a lot of error... And I always try to meet the patient … I think it's better. Because then you physically lay eyes on the patient. And I know it's hard … but at the same time, not looking at the patient happens too much."[19]	Bedside handoffs may provide a different level of safety for learners to practice handoff skills than provider-only locations

ED, emergency department

Table. Continued.

Themes	Representative quotes	Educational considerations
Process		
Provider perception of process and order	"Okay, unless this is just how my mind works as I'm coming onto sign-out, the things that I want to hear .. in this order … up front, is this the sickest patient and/or if this is potentially the sickest patient that I need to worry about because it immediately changes my way of thinking."[17] "[The action list] is kind the heart of the sign out, 'cause it's like 'I couldn't get this done, so you get it done for me' and then we have a disposition."[18] "It's sort of my pet peeve to go out of order in sign outs… I can't follow that."[17]	Importance of standardized process
Provider perception of process, order, collective understanding, and consensus building	"[Standardization is] just predictable, and it's the same every time… I mean, maybe not, you can't do it every time exactly identical, but if we have the same format, then everyone is getting the information they need."[18] "So, when they're going all over the place, I can't really chart in order… it just helps to tell a story, and I think during sign-out a lot of people throw that all out the window and are all over the place."[19] "I want to know what to concentrate on. And when she tells me, 'oh, you have a patient in 5 that's kind of sick and one in 12 that's kind of sick', then I know then to keep an eye on those two and make sure that they're being taken care of."[18] "so the other person doesn't have to find you… so the other person doesn't have to reinvent… what the problem already is."[19] "So making sure that you like spoon feed the critical though process to the next team, and you hope that they rethink it, but they might not, so."[19] "It's [synthesis] is probably really good, because then you know that the person that gave it to you, like, took the key points and was able to kind of succinctly throw it back at you."[19] "So, like, so that I can sort of pick up where they left off, in terms of what was the hold up, what was the problem we saw, and be directed by someone who's been there for twelve hours versus scrambling around trying to find it myself. So having like, you know, where's the first place that you should go."[19] "if there's a vague plan, that has been passed on, that seems that vagueness spirals into badness."[19]	Conceptual understanding of a system as two-fold; the individual provider and the collective group or culture. Information order – influence by cultural expectations, ontological frameworks, and the ED environment – and consensus building through storytelling link these two levels, emphasizing closed-loop communication in educational and assessment processes.

ED, emergency department

needs but individual attending preference, which may lead to a high degree of variability and quality in information communicated. Overall, this nebulous educational format discomfited providers. Unfamiliarity with systems, both existing and new, was a running subtheme that threaded its way through FGs, connected to individuality, fear of change, and cultural expectations and norms that allowed for a high degree of variability. Yet, despite these underlying identified motifs, ED culture was still viewed as a process that could be changed. In fact, providers perceived its potential to evolve, and that educational interventions could be successfully implemented if both individual growth and changing collective expectations were considered. Although interprofessional providers expressed concerns about the current system, they also expressed willingness to adopt new practices and a desire to work together towards a shared goal of improved ED handoff outcomes.

Time

For many participants, time–especially its perceived scarcity–was an integral factor in willingness to embrace a new operational practice. Participants expressed frustration with handoffs perceived to be extraneous or elongated, especially due to interruptions. This desire for brevity also was identified as a potential threat towards provider acceptance of new educational interventions, in that providers may be reticent to buy-in to a new process if it is perceived as too cumbersome or poorly implemented.

Environment

Providers' perceptions of current handoff practices and their willingness to accept future interventions were intricately tied to the environment of the handoff process. Concerns regarding the physical location of the ED sign out were three-fold. First, providers expressed that the lack of an officially designated and consistent location for all handoffs contributed to a poor understanding of when handoff was in progress, and was implicitly linked to frequent and unnecessary interruptions during the handoff process. Participants voiced a perception that a designated handoff location might decrease the frequency of interruptions, and improve patient safety and privacy. Second, although a number of resident and attending participants cited that a non-designated space leads to an increased number of interruptions by ED staff, nurse participants voiced the

experienced reality that isolated areas allow sequestration and a disconnect between interprofessional providers in different roles. The third concern was related to location's effect on provider-patient interactions and relationships. Participants expressed that the current practice of conducting the handoff at a central computer station lacked privacy, not only for participating teams but also for patients. A number of participants advocated for bedside handoffs, as an in-person location was also perceived as providing both effective patient care and a better learning opportunity.

Process

Underlying provider concerns regarding time, location, and culture were perceptions of individual and collective conceptual understanding, and the processes by which these are formed. Conceptual understanding, as espoused throughout the FGs, encompasses how both a single participant receives and synthesizes information (individual conceptual understanding), and how a group interacts with this individual comprehension to collaborate, build consensus, and in turn influence individual thought. Many providers describe this process in terms of order; information is synthesized based on the order in which it is received. Although order was not typically associated with whether or not a system would be implemented successfully, its consistent presence as a vital aspect of the ideal handoff emphasizes the impact this factor has, and how it is impacted in turn by the other identified factors of culture, time, and location.

DISCUSSION

Understanding the three major categories that contribute to the collective conceptual understanding (Figure) of the ED provider handoff has important implications for emergency medicine resident education. The themes identified that create this framework were clearly separate entities in their specifics, yet the categories were also deeply intertwined with each other. Appreciating this interconnectedness while focusing educational interventions to address a learner's understanding of each category and theme will be important for improving resident education in this area of knowledge and practice.

Process was the theme that contributed most significantly to our understanding of the ontological framework, specific tools, and language needed to function within the handoff procedure. Knowledge of clinical emergency medicine vocabulary and the ability to present this information in a format others understood were the two requirements to participate in a handoff. However, in analyzing provider perceptions, it became apparent that the order in which information is given, and thus conceptualized, cannot be divorced from the expectations and environmental factors influencing it. These factors included cultural phenomena (such as individual preferences, individual skill in story-telling, practice and expertise over time); the ontological framework in which the process of handoff is grounded (the assumed knowledge of medical terminology and the language in which action–tests, labs, prognosis, etc.,–is couched in); and the environmental realities of working in a fast-paced, high-volume setting. These factors form a collective conceptual understanding of the general ED setting that influence the type of information passed along and in what manner. An individual giving handoff draws from this information set in order to pass along information to the oncoming team and thus help form individual conceptual understanding. Therefore, providers expressing frustration with lack of order are actively calling upon, and critiquing, this foundational background that shapes how the information is being relayed to them. The other themes identified from provider perceptions inform how this process of collaboration and consensus building plays out and to what extent it is efficacious.

The factors identified in the Environment theme were consistent with previously reported factors that impact ED handoffs (location, interruptions, ambient noise level, etc).[1,7] While the physical layout of the ED may make some of the environmental factors more or less of an issue for handoffs, we did not identify any factors that were not previously reported or unique only to this academic ED. Still, handoff location was perceived as not only the background for ED handoff culture, but also a direct influence on how providers interacted with each other within that culture. Handoff location plays a central role, both positive and negative, in interprofessional relationships, as the site of both collaboration and relationship formation. Additionally, location, and perceptions of duties in the context of specific locations, influences the efficacy of collaboration and consensus building between providers. This can be seen in the perception of interprofessional interactions between nursing and resident/attending providers; depending on the location, these moments have the potential to disrupt collaboration (i.e., interruptions), or allow for further consensus building and learning. Cultural expectations may potentially evolve at locations that bring together patients and interprofessional providers. Location is not simply a utilitarian factor, a shield from the surrounding chaos, but also a potential barrier, or facilitator, of effective education and assessment.

The themes of Culture and Time, and subthemes of cultural expectations and accepted practices identified by our team all have the potential to impact Process. Individuality, hierarchies and differences in expectations between various roles participating in the handoff were significant subthemes identified, and educational interventions designed to teach and assess handoff performance can avoid related barriers by defining roles and setting standard expectations. Additionally, the relationship between culture and location, explored above, was an important interprofessional subtheme identified. Although the "ideal" setting for handoffs is unknown, assessing interprofessional perceptions of handoff locations that best facilitate collaboration is an important precursor to implementing handoff educational interventions. Finally, many of the environmental barriers to optimal ED sign-out may only be minimized, rather than eliminated. Likewise, the

ontological framework and language in which educational interventions take place may be difficult to alter due to the larger culture of the medical system. Educators designing interventions to teach, assess, and ultimately improve ED handoffs might prioritize efforts to change the underlying culture, as the effectiveness of these interventions may rely less on specific procedural changes and more on how they change cultural expectations, perceptions, and norms.

LIMITATIONS

A potential limitation of this qualitative study is that we performed purposive sampling from a single center. However, we asked participants to rely upon their experiences at all prior settings when answering questions regarding handoff perceptions, and therefore believe the perceptions represent multiple ED settings. Although efforts were made to ensure thematic saturation and data credibility, it is possible there are additional relevant themes that were not uncovered by our study. Finally, although the sampling and FG structure were purposive to facilitate interprofessional discussion, it is possible additional themes would have been uncovered if groups were stratified by discipline.

CONCLUSION

Interprofessional ED providers in this qualitative study identified four major categories that contribute to the collective conceptual understanding of the ED provider handoff. Understanding this framework and the themes that create it has important implications for emergency medicine resident education. Educators wishing to develop educational interventions to improve resident education in emergency medicine handoff knowledge and practice may wish to explicitly consider how the intervention may impact and interact with these factors, as they may affect learners' acceptance and incorporation of the intervention.

Address for Correspondence: James A. Heilman, MD, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd. Portland, OR 97239-3079. Email: heilman@ohsu.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Arora V and Johnson J. A Model for Building a Standardized Handoff Protocol. *Jt Comm J Qual Patient Saf.* 2006;32(11):646-55.
2. Accreditation Council for Graduate Medical Education. Resident Duty Hours and the Working Environment: Comparisons of 2003 and 2011 Standards. Available at: http://www.acgme.org/acgmeweb/Portals/0/PDFs/dh-ComparisonTable2003v2011.pdf. Accessed Mar 2015.
3. Starmer AJ, Spector ND, Srivastava R, et al. Changes in Medical Errors after Implementation of a Handoff Program. *N Engl J Med.* 2014;371(19):1803-12.
4. Arora VM, Greenstein EA, Woodruff JN, et al. Implementing Peer Evaluation of Handoffs: Associations with Experience and Workload. *J Hosp Med.* 2012;8(3):132-6.
5. Cheung DS, Kelly JJ, Beach C, et al. Improving Handoffs in the Emergency Department. *Ann Emerg Med.* 2010;55(2):171-80.
6. Gibson SC, Ham JJ, Apker J, et al. Communication, Communication, Communication: The Art of the Handoff. *Ann Emerg Med.* 2010;55(2):181-3.
7. Kessler C, Shakeel F, Hern HG, et al. An Algorithm for Transition of Care in the Emergency Department. *Acad Emerg Med.* 2013;20(6):605-9.
8. Petersen SM, Gurses AP, Regan L. Resident to Resident Handoffs in the Emergency Department: An Observational Study. *J Emerg Med.* 2014;47(5):573-79.
9. Starmer AJ, Sectish TC, Simon DW, et al. Rates of Medical Errors and Preventable Adverse Events Among Hospitalized Children Following Implementation of a Resident Handoff Bundle. *JAMA.* 2013;310(21):2262-70.
10. Starmer AJ, Spector ND, Srivastava R, et al. I-PASS, a Mnemonic to Standardize Verbal Handoffs. *Pediatrics.* 2012;129(2):201-4.
11. Sujan MA and Spurgeon P. Safety of Patient Handover in Emergency Care – Results of a Qualitative Study. Proceedings of Annual European Safety & Reliability Conference. 2013.
12. Kern D, Thomas P, Hughes M. Curriculum Development for Medical Education – A Six-Step Approach. Second Edition. The Johns Hopkins University Press. 2009.
13. Watling CJ and Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Med Teach*. 2012;34(10):850-61.
14. Sullivan G and Sargeant J. Qualities of Qualitative Research: Part 1. *J Grad Med Educ.* 2011;3(4):449-52.
15. Sargeant J. Qualitative Research Part II: Participants, Analysis, and Quality Assurance. *J Grad Med Educ.* 2012;4(1):1-3.
16. Charmaz K. Constructing Grounded Theory – A Practical Guide Through Qualitative Analysis. SAGE Pulbications Ltd. 2010.
17. MF. Focus group transcript. 2014, October 8th.
18. MF. Focus group transcript. 2014, October 14th.
19. MF. Focus group transcript. 2014, October 21st.

The Impact of Medical Student Participation in Emergency Medicine Patient Care on Departmental Press Ganey Scores

Aaron W. Bernard, MD*
Daniel R. Martin, MD†
Mark G. Moseley, MD†
Nicholas E. Kman, MD†
Sorabh Khandelwal, MD†
Daniel Carpenter, MPH‡
David P. Way, MEd†
Jeffrey M. Caterino, MD, MPH†

*Quinnipiac University, Frank H. Netter MD School of Medicine, Hamden, Connecticut
†Ohio State University College of Medicine, Department of Emergency Medicine, Columbus, Ohio
‡Ohio State University, Department of Biomedical Informatics, Columbus, Ohio

Section Editor: David Wald, DO
Submission history: Submitted May 1, 2015; Revision received September 16, 2015; Accepted September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27321

Introduction: Press Ganey (PG) scores are used by public entities to gauge the quality of patient care from medical facilities in the United States. Academic health centers (AHCs) are charged with educating the new generation of doctors, but rely heavily on PG scores for their business operation. AHCs need to know what impact medical student involvement has on patient care and their PG scores.

Purpose: We sought to identify the impact students have on emergency department (ED) PG scores related to overall visit and the treating physician's performance.

Methods: This was a retrospective, observational cohort study of discharged ED patients who completed PG satisfaction surveys at one academic, and one community-based ED. Outcomes were responses to questions about the overall visit assessment and doctor's care, measured on a five-point scale. We compared the distribution of responses for each question through proportions with 95% confidence intervals (CIs) stratified by medical student participation. For each question, we constructed a multivariable ordinal logistic regression model including medical student involvement and other independent variables known to affect PG scores.

Results: We analyzed 2,753 encounters, of which 259 (9.4%) had medical student involvement. For all questions, there were no appreciable differences in patient responses when stratifying by medical student involvement. In regression models, medical student involvement was not associated with PG score for any outcome, including overall rating of care (odds ratio [OR] 1.10, 95% CI [0.90-1.34]) or likelihood of recommending our EDs (OR 1.07, 95% CI [0.86-1.32]). Findings were similar when each ED was analyzed individually.

Conclusion: We found that medical student involvement in patient care did not adversely impact ED PG scores in discharged patients. Neither overall scores nor physician-specific scores were impacted. Results were similar at both the academic medical center and the community teaching hospital at our institution. [West J Emerg Med. 2015;16(6):830–838.]

INTRODUCTION

Press Ganey (PG) scores are an important marker of quality medical care, used by hospital administrators, healthcare consumers, and payers.[1-3] Subsequently, reimbursements are being linked to these measurements.[4] Researchers have identified variables associated with PG scores including interpersonal interactions, patient communication, and perceived wait time.[1-4] Other factors such as age, race/ethnicity, triage acuity, and arrival time have also been suggested to affect PG scores.[1-5]

The impact of medical student involvement on an emergency departments (ED's) PG scores is not well defined in the literature. A convenience sample of 145 patients in an ED located in Ireland suggested positive patient attitudes towards medical students.[6] Similar studies of non-ED ambulatory settings have also reported positive patient opinions about medical students.[7-9]

ED student rotations have been increasing over the last decade.[10,11] Medical school enrollment is also increasing, requiring additional clinical teaching sites to accommodate demand.[12] While reports from other settings provide reassurance that medical students are well received by patients, ED studies are limited.[6-9] Furthermore, recent literature indicates that patients have difficulty distinguishing between various providers.[13,14] Accordingly, there is a possibility that students could not only impact an ED's overall PG scores, but also the scores of physician providers. The potential for negative student impacts on PG scores could hinder developing partnerships between EDs and medical schools, and could ultimately harm provider reimbursement. Therefore, a better understanding of the relationship between medical students and EDs' PG measures is needed. The goal of this investigation was to determine whether medical student involvement in emergency care impacts our ED's PG scores.

METHODS

Study Design and Setting

We conducted a retrospective, observational cohort study examining the relationship between medical student involvement in ED care and PG survey scores, a common surrogate measure of patient satisfaction. The study was approved by the hospital institutional review board and was conducted in compliance with the Strengthening the Reporting of Observational studies in Epidemiology statement.[15]

The Ohio State University medical facility provides patient care in two EDs. The first is an academic, tertiary care, Level I trauma center with a volume of 72,000 patient visits per year (the "academic ED"). The academic ED patient population is diverse with regard to ethnicity/race and economic status. The second is a community teaching ED with a volume of 50,000 patient visits per year (the "community ED"). The community ED patient population is primarily African-American and of lower economic status. Both sites are staffed by the same physician group. Staffing consists of board-certified emergency physicians, emergency medicine resident physicians, and rotating resident physicians. The academic ED averages seven residents per shift and the majority of patients have a resident involved in their care. Advanced-Practice providers (APPs; nurse practitioners and physician assistants) primarily care for lower acuity patients. The community ED typically has either one resident or one APP on each shift, with additional APP coverage for lower acuity patients.

The Ohio State University College of Medicine enrolls 210 students per year. All fourth-year medical students take a required, four-week ED clerkship at one of 10 EDs in the Central Ohio area. About 95, (45%) are assigned to the two study EDs, averaging about eight per month. Additionally, the study EDs take two third-year elective students per month.

Medical students select patients who have been identified in the electronic medical record (EMR) as having been triaged and placed in a room, i.e. "waiting for provider." Once a patient is selected, the student performs a focused history and physical examination, and then presents the case to the supervising physician. Students are encouraged to select patients with a broad range of chief complaints and triage acuity. Critically ill patients are rarely labeled as "waiting for provider," and therefore students are not involved with their care. Students are not provided any specific education about PG surveys or patient satisfaction.

Selection of Participants

All patients seen at either the Ohio State University Main Hospital ED or their affiliate community hospital ED and who completed a PG survey for visits between December 1, 2011, and December 31, 2012, were eligible for the study. Eligible patients receive a survey if they were age 18 or over and discharged from the ED. Patients admitted or placed in observation units were not eligible and were excluded from the study. Patients were also excluded if they failed to answer both primary outcome items: H.1 "Overall rating of care received during your visit" and H.2 "Likelihood of your recommending our ED to others." Exclusion criteria also included lack of attending physician involvement in care or missing documentation about attending providers.

Measurement

Adult patients discharged from the EDs are randomly selected by PG and receive a telephone survey call. Trained PG personnel administer surveys. Five call attempts are made at different times for each patient selected. Additional patients are called until a quota of 230 completed surveys per month for both EDs combined is reached. PG does not report traditional response rates for phone surveys, but they estimate a rate of one completed survey for every eight patients called.

The PG ED satisfaction survey consists of 31 questions

organized into eight sections: arrival, nurses, doctors, tests, family or friends, personal/insurance information, personal issues, and overall assessment. PG reports rigorous psychometric testing as part of survey design and evaluation to provide reliable and valid data.[16]

Data for this study was electronically obtained from the institution's central data repository. Clinical data is completed using the EPIC (EPIC, Verona, WI) EMR. Variables included patient and visit characteristics as well as results of the PG survey. We performed range checking for all variables.

The ED visited (academic or community ED) was identified. Abstracted patient characteristics included age and race/ethnicity, which was coded as White, Hispanic, Black, or other. Hispanics were combined with Whites for analyses because of the small number of Hispanic patients resulting in complete or quasi-complete separation in the multivariable ordinal models.

We categorized month of visit by season: winter (January–March), spring (April-June), summer (July–September), or fall (October–December). Arrival time was categorized by shift: day (7am-3pm), evening (3pm-11pm), or night (11pm-7am). Emergency severity index (ESI) at triage was calculated by nursing staff for all ED patients.[17] Given a limited number of Level I encounters (highest acuity patients often by-pass the ED) and level 5 encounters, the ESI was coded as three levels: ESI 1-2, ESI 3, and ESI 4-5. Primary payer type was coded as managed care, private insurance, Medicaid, Medicare, other governmental payer, or self-pay. We identified ED length of stay as time of arrival until time of discharge. Because of outliers and non-linear distribution, length of stay variable was divided into quartiles. We created dichotomous variables to note the use of plain radiographs, computed tomography (CT), and laboratory tests for each visit.

Provider variables included whether a medical student or resident physician participated in the patient care, and the attending physician who discharged the patient. We excluded patients with missing responses to a specific PG question, from analysis of that question.

Outcomes

Outcome variables were patient responses to PG questions related to overall satisfaction and to satisfaction with physician care. All responses were ordinal variables scored on a five-point scale with 1 being very poor and 5 being very good. The items of primary interest were overall rating of ED care (H.1) and likelihood of recommending ED to others (H.2). Of secondary interest were four items regarding physician behavior (C.1-C.4).

Analysis

Descriptive data included proportions, means with standard deviations (SD), and medians with interquartile range (IQR) as appropriate. Normality was tested using the Shapiro-Wilk test. We made comparisons of patient responses between those who experienced medical student participation and those who did not using chi-square tests. Analyses were performed with STATA v12 (STATACorp, College Station, TX).

We first reported the distribution of patient responses to each of the study questions stratified by medical student participation in the patient's care. We next constructed ordinal logistic regression models clustered by attending physician for each question. After constructing an initial unadjusted univariate model including only medical student involvement and the clustering, we then constructed adjusted multivariable models for each outcome. Independent variables were chosen a priori from factors previously shown to affect patient satisfaction scores.[1-5] These included medical student involvement, ED visited, age, race, season of visit, time of arrival, ESI level, payer type, ED length of stay, ordering of ≥1 radiograph, ≥1 CT, ≥1 laboratory test, and resident involvement. The highest level of recommendation (5–very good) was used as the reference group. Resulting odds ratios (OR) >1 indicate greater odds of having lower satisfaction scores. In each model we tested for interactions between medical student involvement and: patient age, ESI level, and resident involvement.

Variables in each model were tested for violation of the proportional hazards (parallel lines) assumption using the Brant test. A partial proportional odds model was then created using the STATA gologit2 command to allow variables violating the assumption to vary across response levels.[18,19] Variables not violating the proportional odds assumption continued to be held constant across response levels. To determine the best fitting model for each question, we calculated the Akaike Information Criterion (AIC) for each of the models created for that question.[20] As a sensitivity analysis, we analyzed each question for each of the two EDs individually. We also performed a sensitivity analysis adding a satisfaction with nursing care variable to the model. This was based on the average score of each patient in the nursing-related questions of the PG survey.

Sample size requirements for multivariable ordinal logistic regression are not clearly defined.[21] The rule of thumb for logistic regression is 10 outcome events per independent variable. In 13 months, we expected to have 2,990 surveys for analysis and we expected that 20% (n=598) of our subjects would report satisfaction scores of <5. This would provide 20 patients with scores <5 (i.e., 20 outcome events) for each of up to 30 independent variables (accounting for multilevel nature of several categorical variables).

Finally, because PG surveys are completed by a minority of the discharged ED population, we obtained descriptive statistics on the entire discharged ED population during the study time period to identify differences in characteristics between patients completing the survey and those who did not.

RESULTS

Initially, 3,421 ED encounters with returned PG surveys

were identified. Exclusion criteria eliminated 668, leaving 2,753 surveys for the study: (353 were missing answers to the primary outcome questions, 295 lacked attending involvement, 14 were missing documentation of specific attending involved, and 6 were admitted/observation patients; see Figure 1 for a flow chart description). Missing outcome data included a lack of patient response to questions C.1-C.4: 52 missing for C1, 56 for C.2, 66 for C.3, and 67 for C.4. Forty-five patients did not respond to any of the four questions. We retained all patients for the analysis of the primary outcomes (questions H.1 and H.2), but those with missing responses to specific questions C.1-C.4 were excluded from analysis of that question. Although our initial expectations were to have 2,990 surveys for the study, for the primary analysis there were greater than 40 outcome events (score less than 5 or "very good") per independent variable or variable level in each model. This is well above the recommended 10 outcome events per independent variable.[21]

Population demographics are shown in v 1. For the race/ethnicity variable, the White or Hispanic category included 52 Hispanics and the other category included 50 Asians and 5 Native Americans. There were 42 attending physicians who worked shifts during the study period (December 1, 2011 through December 31, 2012). The 42 physicians encountered an average of 65 patients who were PG respondents during that time (Mean=65; SD= 38). Attendings encountered an average of six PG respondent patients with the involvement of a medical student (Mean=6; SD=4).

There were 128 medical students who rotated in the university EDs during the study period (104 fourth-year students, 24 third-year students). Of the 2,753 encounters analyzed, 259 (9.4%) had medical student involvement. Resident physicians were involved in 52% of cases. In the academic ED, 71% of patients had resident involvement in their care and 10% had medical student involvement. In the community ED, 19% of patients had resident involvement in their care and 7.4% had medical student involvement. Characteristics of patients were similar between those with and without medical student involvement (Table 1). Patients with medical student involvement were less likely to also have resident involvement (32% versus 54%). Length of stay was longer for patients with medical students.

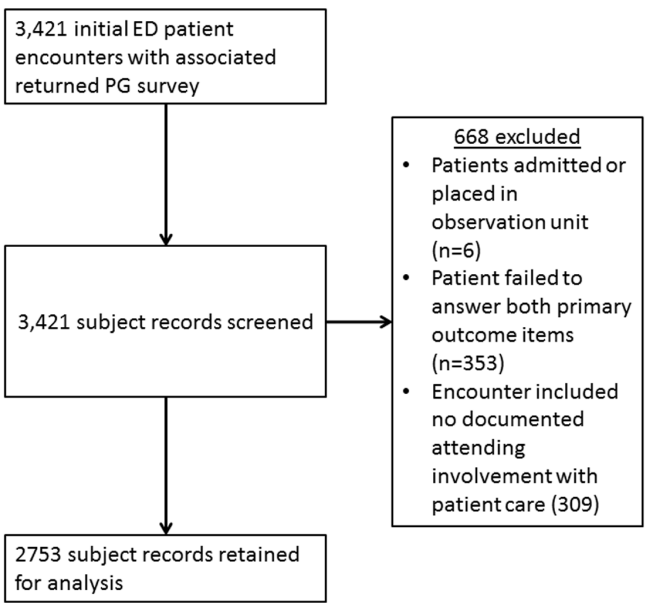


Figure 1. Flow chart describing identification of patient subjects for study.
ED, emergency department; *PG,* press ganey

The distribution of responses was similar between encounters with and without medical student involvement for all satisfaction questions. The chi-square tests resulted in p values greater than 0.05 for all PG questions comparing medical student participation to non-medical student participation (see Figure 2 and Figure 3). Most ratings (>80%) were good or very good for each question.

In the ordinal regression models, medical student involvement was not significantly associated with PG score for any outcome measure, either in the adjusted or unadjusted analyses. Figure 4 notes which variables violated the proportional odds assumption in each model. There were no significant interactions in any model. Fit was equivalent between models constructed for each question with no appreciable differences in AIC between the models (data not shown). The CIs for medical student involvement were not appreciably wider in the adjusted versus the unadjusted model, indicating that there was likely no instability caused by the number of independent variables in the models.

In the sensitivity analysis, the effect of medical student involvement remained non-significant for all questions when analyzed at each ED site. For question H1 (overall rating of ED care), the results OR for medical student involvement at the academic ED was 1.08 (95% CI [0.81-1.46]) (p=0.587) and for the community ED was 1.11 (95% CI [0.66-1.87]) (p=0.689). For question H2 (likelihood of recommending), the results OR for medical student involvement at the academic ED was 1.02 (95% CI [0.74-1.40]) (p=0.892) and for the community ED was 1.20 (95% CI [0.77-1.87]) (p=0.413).

Table 2 shows descriptive values for the study population compared to the entire discharged ED population. Patients seen in the academic ED, were more likely to be White, with lower ESIs, shorter lengths of stay, and more likely to have private insurance. Other characteristics were similar between EDs, particularly, rates of medical student and resident involvement.

DISCUSSION

This investigation provides evidence that medical student involvement in emergency care does not adversely impact PG scores. Neither overall ED scores nor the physician provider scores were impacted by medical student involvement. Results were similar across both an academic medical center and an affiliated community teaching hospital.

Table 1. Percentage and number (in parentheses) of 2,753 discharged emergency department (ED) patients with completed Press Ganey survey results for the total population and broken down by medical student vs. no medical student participation separately.

Variable	Total (n=2,753)	Medical student involved (n=259)	No medical student involved (n=2,494)
ED visited			
Academic	64% (1,773)	72% (186)	64% (1,587)
Community	36 (980)	28 (73)	37 (907)
Age (in years)*	38 (26-54)	40 (27-55)	38 (25-54)
Race			
African American or Black	38 (1,057)	36 (93)	39 (964)
White or Hispanic	58 (1,588)	61 (159)	55 (1,375)
Other	3.9 (108)	2.7 (7)	6.2 (155)
Season of visit			
Winter (Jan-Mar)	34 (922)	35 (90)	33 (9,832)
Spring (Apr-June)	22 (598)	12 (31)	23 (567)
Summer (July-Sept)	23 (644)	38 (99)	22 (545)
Fall (Oct-Dec)	21 (589)	15 (39)	22 (550)
Time of arrival			
Day shift (7am-3pm)	44 (1,211)	54 (139)	43 (1,072)
Evening shift (3pm-11pm)	40 (1,107)	37 (96)	41 (1,011)
Night shift (11pm-7am)	16 (435)	9.3 (24)	16 (411)
Emergency severity index			
Level 1 & 2	13 (354)	12 (31)	13 (323)
Level 3	54 (1,490)	57 (148)	54 (1,342)
Level 4 & 5	33 (909)	31 (80)	33 (829)
Payer type			
Managed care/private insurance	35 (975)	34 (88)	36 (887)
Medicaid	27 (731)	29 (76)	26 (655)
Medicare	17 (465)	16 (42)	17 (423)
Other government payer	2.7 (75)	3.5 (9)	2.6 (66)
Self pay	18 (507)	17 (44)	19 (463)
ED length of stay (in minutes)*	227 (142-338)	249 (161-355)	224 (141-335)
X-ray was ordered	32 (887)	30 (77)	32 (810)
CT was ordered	15 (419)	18 (46)	15 (373)
Lab tests were ordered	59 (1,613)	61 (158)	58 (1,455)
Providers			
Medical student involved in care	9.4 (259)	-	-
Resident physician involved in care	52 (1,442)	32 (84)	54 (1,358)

CT, computed tomography
*Medians and interquartile ranges (IQR) are provided for Age and ED Length of Stay.

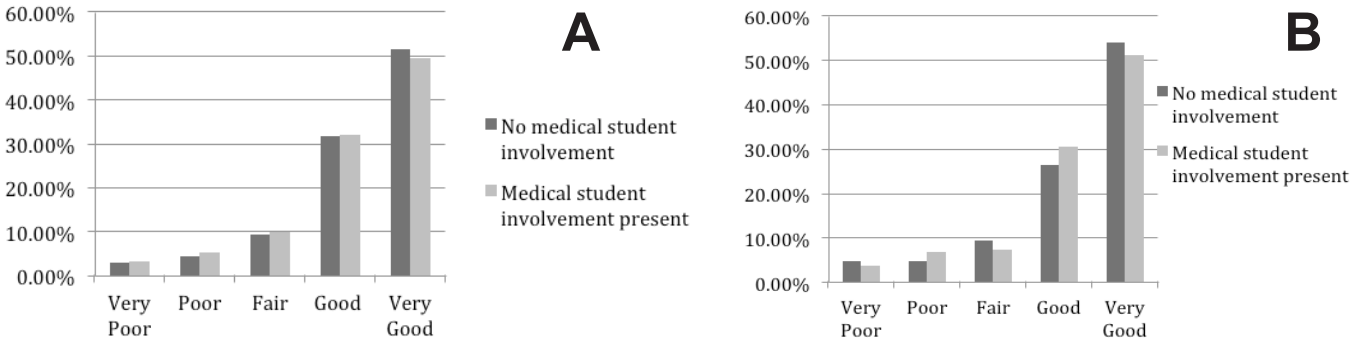


Figure 2. Patient response to primary outcome questions described graphically as a percentage of responses. A, Overall rating of care received during your visit. B, Likelihood of your recommending our emergency department to others.

		MS Involved					No MS Involved					
	N	VP	P	F	G	VG	VP	P	F	G	VG	p
H1. Overall rating of your care received during your visit	2753	3.1(8)	5.4(14)	10(26)	32(83)	49(128)	3.1(76)	4.4(110)	9.3(231)	32(794)	51(1283)	0.93
H2. Likelihood of your recommending our ED to others	2753	3.9(10)	7.0(18)	7.3(19)	31(79)	51(133)	4.8(120)	4.9(123)	9.5(238)	27(663)	54(1350)	0.27
C1. Courtesy of the Doctor	2701	1.9(5)	4.3(11)	7.7(20)	28(73)	57(147)	1.3(32)	2.8(70)	5.7(141)	29(724)	59(1,478)	0.36
C2. Degree to which the doctor took the time to listen to you	2697	2.7(7)	5.0(13)	6.6(17)	28(72)	57(147)	2.1(52)	4.1(102)	7.5(188)	28(700)	56(1,399)	0.88
C3. Doctor's concern to keep you informed about your treatment	2687	3.1(8)	6.0(15)	11(29)	25(66)	54(139)	2.2(55)	6.0(149)	9.3(232)	27(669)	53(1,325)	0.78
C4. Doctor's concern for your comfort while treating you	2686	2.7(7)	5.4(14)	8.5(22)	29(74)	53(138)	2.7(67)	4.5(112)	8.3(207)	29(723)	53(1,322)	0.98

Figure 3. Percentage and number (in parentheses) of patient responses to Press Ganey survey questions about their care by an emergency department stratified by medical student involvement in their care.
MS, medical student; *N*, number of patient respondents; *VP*, very poor; *P*, poor; *F*, fair; *G*, good; *VG*, very good; *ED*, emergency department

Question	Unadjusted odds ratio for medical student involvement*			Adjusted odds ratio for medical student†			Variables violating the proportional odds assumption‡
	OR	95% CI	p	OR	95% CI	p	
H.1	1.10	(0.90-1.34)	0.37	1.12	(0.88-1.42)	0.36	Race, payer
H.2	1.07	(0.86-1.32)	0.54	1.09	(0.86-1.38)	0.48	Race, payer, department, arrival shift, ≥1 x-ray obtained
C.1	1.19	(0.96-1.48)	0.12	1.20	(0.96-1.48)	0.10	Payer
C.2	1.01	(0.82-1.25)	0.92	1.01	(0.79-1.28)	0.95	Race, payer, age, resident involvement
C.3	1.05	(0.85-1.30)	0.64	1.05	(0.84-1.30)	0.67	Race, payer
C.4	1.03	(0.83-1.27)	0.81	1.01	(0.80-1.27)	0.93	Race, payer, ESI at triage

Figure 4. Results of ordinal logistic regression models identifying the effect of medical student involvement in emergency department care in causing decreased patient satisfaction scores.
OR, odds ratio; *CI,* confidence interval; *ESI*, emergency severity index
*Using a univariate ordinal logistic regression model clustered by attending.
†Using a partial proportional odds ordinal logistic regression model clustered by attending physician and controlling for: medical student involvement, age, race, department, resident involvement, ESI at triage, primary payer, arrival shift, season of visit, length of stay, ≥1 x-ray obtained, ≥1 computed tomography obtained, and ≥1 laboratory study obtained.
‡Violations accounted for in the partial proportional odds models.

To the best of our knowledge, this was the first ED study to consider the impact of medical students on PG measures. This distinction is important as PG surveys are generally considered the benchmark for satisfaction goals and may directly impact an institution's or physician provider's reputation and financial reimbursement.[4]

The American College of Emergency Physicians (ACEP) recognizes the value of patient satisfaction surveys and makes recommendation regarding optimal features of survey tools.[22] PG methodology is compliant with most ACEP

Table 2. Percentage of 2,753 discharged emergency department (ED) patients with completed Press Ganey (PG) survey results compared to the entire population of patients discharged from the study EDs.

Variable	Patients with PG surveys (n=2,753)	All patients discharged from study EDs (n=111,180)
ED visited		
Academic	64%	55%
Community	36	45
Age (in years)*	38	42 (28-54)
Race		
African American or Black	38	47
White or Hispanic	58	48
Other	3.9	5.0
Season of visit		
Winter (Jan-Mar)	34	30
Spring (Apr-June)	22	24
Summer (July-Sept)	23	24
Fall (Oct-Dec)	21	23
Time of arrival		
Day shift (7am-3pm)	44	37
Evening shift (3pm-11pm)	40	41
Night shift (11pm-7am)	16	21
Emergency severity index (ESI)†		
Level 1 & 2	13	21
Level 3	54	46
Level 4 & 5	33	30
Payer type		
Managed care/private insurance	35	24
Medicaid	27	30
Medicare	17	19
Other government payer	2.7	2.0
Self pay	18	24
ED length of stay (in minutes)*	227 (142-338)	298 (108-347)
X-ray was ordered	32	38
Computed tomography was ordered	15	18
Lab tests were ordered	59	59
Providers		
Medical student involved in care	9.4	8.6
Resident physician involved in care	52	48

*Medians and interquartile ranges (IQR) are provided for Age and ED Length of Stay.
†ESI measurement not available for 2.2% of the entire ED population.

recommendations including transparency of process and analysis, consideration for education level survey subjects, administration close to service date, and collection of discrete data points.[16] The basis of criticism of most surveys, including PG, involves the ACEP recommendations that surveys have a statistically valid sample size, and are free from selection bias.[23,24] Despite this limitation, PG is the most commonly used patient satisfaction survey service, and represents the standard across the healthcare industry.[16] The broad use of PG surveys and comparability across settings, has prompted similar ED patient satisfaction research.[25-28]

Our results should provide reassurance to clerkship directors, medical directors, and hospital administrators that student education does not come at the expense of PG scores. This is particularly important since emergency medicine is being increasingly recognized as an important learning experience for medical students.[10,11] Furthermore, this conclusion extends across two very different ED settings.

While most similar research has resulted in improved patient satisfaction from medical student involvement, our findings were neutral. The design of our study does not explain why we found a neutral result. One potential reason may be that the medical student effect is too small to detect in one full year's worth of patient data. Another potential reason may be the difference in care setting. Patients have reported apprehension regarding student involvement in intimate exams, in cases that are more emotional, and in more serious situations. These are all common occurrences in a busy ED.[6-9] Another potential confounding variable could be longer lengths of stay associated with medical student involvement.

We found that length of stay increased by 25 minutes, a finding consistent with other studies.[29] Further work is needed to explain the impact of medical student involvement on patient satisfaction.

Engaging medical students in the PG survey could improve PG scores. Providing students with patient satisfaction skills might promote better interpersonal interactions and better patient communications.[1-4] Perhaps coaching medical students to more frequently check on patient's needs could mitigate the longer stays seen with student involvement.[4]

LIMITATIONS

Our results represent the experience of one institution with an established history of clinical teaching, one group of physicians, and medical students from one medical school. However, we included an academic and a traditional community ED, each with very different staffing models and patient populations. We hope this supports the generalizability of our findings. Although we accounted for clustering at the attending level we did not explore additional levels of clustering, such as specific attending-medical student dyads.

There are several limitations inherent to the PG phone methodology including limited language options, need for patients to own a phone, and an inability to identify true survey response rates. We compensated by trying to compare the survey population to the population of discharged ED patients as a whole. Based on the characteristics of the two study EDs, we believe that the greater rates of Whites and private insurance in the study population were driven by the greater proportion of visits to the academic ED as compared to the overall discharged ED population. Our community ED's population is more likely to be Black and self-pay. The results of the sensitivity analysis which was limited to the community ED and which were consistent with our overall results provides reassurance that these differences did not cause biased results. Study patients appeared to have shorter length of stay than the discharged population as a whole. However, it is unclear how this may have impacted our outcomes.

We were able to abstract most relevant variables from our EMR except door-to-doctor time, which may have resulted in a degree of unmeasured confounding. There also may be unaccounted differences between third- and fourth-year students, including interest in the rotation and clinical skill. The impact of these potential differences on PG scores is unclear.

CONCLUSIONS

We found that medical student involvement in ED care does not adversely impact PG scores. Neither overall scores nor physician scores were impacted by medical student involvement at our institution. Further, the results were similar across both an academic medical center and the community teaching hospital.

ACKNOWLEDGMENTS

The authors would like to thank Luke Macadam for his assistance with this project.

Address for Correspondence: Aaron W. Bernard, MD, Frank H. Netter MD School of Medicine at Quinnipiac University, 275 Mt Carmel Ave, Hamden, CT 06518-1908. Email: aaron.bernard@quinnipiac.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Boudreaux ED and O'Hea EL. Patient satisfaction in the emergency department: a review of the literature and implications for practice. *J Emer Med*. 2004;26:13-26.
2. Brown AD, Sandoval GA, Levinton C, et al. Developing an efficient model to select emergency department satisfaction improvement strategies. *Ann Emerg Med*. 2005;46:3-10.
3. Taylor C and Benger JR. Patient satisfaction in emergency medicine. *Emerg Med J*. 2004;21:528-32.
4. 2011 Press Ganey Pulse Report - Perspectives on American health care. Available at www.pressganey.com. Accessed Dec 15, 2013.
5. Sun BC, Adams J, Orav EJ, et al. Determinants of patient satisfaction and willingness to return with emergency care. *Ann Emerg Med*. 2000;35:426-534.
6. Kuan S and O'Donnell JJ. Medical students in the emergency department: how do patients view participation in clinical teaching? *Ir Med J*. 2007;100:560-1.
7. O'Malley PG, Omori DM, Landry FJ, et al. A prospective study to assess the effect of ambulatory teaching on patient satisfaction. *Acad Med*. 1997;72(11):1015-7.
8. Haffling AC and Hakansson A. Patients consulting with students in general practice: survey of patients' satisfaction and their role in teaching. *Med Teach.* 2008;30:622-9.
9. Sperry J, Oreskovich J, Carlisle R, et al. Patient satisfaction with medical student participation in a family practice clinic. *Ann Behav Science in Med Educ*. 2011;17:14-5.
10. Wald DA, Manthey DE, Kruus L, et al. The state of the clerkship: a survey of emergency medicine clerkship directors. *Acad Emerg Med*. 2007;14:629-34.
11. Mulcare MR, Suh EH, Tews M, et al. Third-year medical student rotations in emergency medicine: a survey of current practices. *Acad Emerg Med*. 2011;18:S41-7.
12. Results of the 2012 medical school enrollment study. Available at www.aamc.org. Association of American Medical Colleges.

Washington, DC. Accessed Dec 10, 2013.
13. Olson DP and Windish DM. Communication discrepancies between physicians and hospitalized patients. *Arch Intern Med*. 2010;170:1302-7.
14. Arora V, Gangireddy S, Mehrotra A, et al. Ability of hospitalized patients to identify their in-hospital physicians. *Arch Intern Med.* 2009;162:199-201.
15. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61(4):344-9.
16. Press Ganey Associates, Inc. Available at: www.pressganey.com. Accessed Dec 10, 2013.
17. Emergency severity index (ESI): a triage tool for emergency departments. Version 4. Available at: www.ahrq.gov. Accessed Dec 15, 2013.
18. Williams R. Generalized ordered logit / partial proportional odds models for ordinal dependent variables. *Stata J.* 2006;6:58-82.
19. Gologit2documentation. Available at: http://www3.nd.edu/~rwilliam/stata/gologit2.pdf. Accessed Dec 10, 2013.
20. Hardin JW and Hilbe JM. *Generalized linear models and extensions*. Third edition. College Station, TX. 2012.
21. Stata data analysis examples: Ordered logistic regression. UCLA: Statistical Consulting Group. Available at: http://www.ats.ucla.edu/stat/stata/dae/ologit.htm. Accessed Dec 15, 2013.
22. Patient satisfaction surveys. Policy statement. *Ann Emerg Med*. 2011; 57(3):313.
23. Wears RL. Patient satisfaction and the curse of kelvin. *Ann Emerg Med*. 2005;46:11-2.
24. Seven things you may not know about Press Ganey Statistics. Available at www.epmonthly.com. Accessed Dec 15, 2013.
25. Pines JM, Iyer S, Disbot M. The effect of emergency department crowding on patient satisfaction for admitted patients. *Ann Emerg Med*. 2008;15:825-31.
26. Tekwani KL, Karem Y, Mistry CD. Emergency department crowding is associated with reduced satisfaction scores in patients discharged from the emergency department. *West J Emerg Med*. 2013;14(1):11-5.
27. DeBehnke D and Decker C. The effects of physician-nurse patient care team on patient satisfaction in an academic ED. *Am J Emerg Med*. 2002;20:267-70.
28. Katz DA, Aufderheide TP, Gaeth G, et al. Satisfaction and emergency department revisits in patients with possible acute coronary syndrome. *J Emerg Med*. 2013;45(6):947-57.
29. DeLaney M, Zimmerman KD, Strout TD, et al. The effect of medical students and residents on measures of effiency and timeliness in an academic medical center emergency department. *Acad Med*. 2013;88:172331.

What is the Prevalence and Success of Remediation of Emergency Medicine Residents?

Mark Silverberg, MD*
Moshe Weizberg, MD†
Tiffany Murano, MD‡
Jessica L. Smith, MD§
John C. Burkhardt, MD, MA¶
Sally A. Santen, MD, PhD¶

*SUNY Downstate/Kings County Hospital, Department of Emergency Medicine, New York, New York
†Staten Island University Hospital, Department of Emergency Medicine, Staten Island, New York
‡Rutgers New Jersey Medical School, Department of Emergency Medicine, Newark, New Jersey
§Alpert Medical School of Brown University, Department of Emergency Medicine, Providence, Rhode Island
¶University of Michigan Medical School, Department of Emergency Medicine, Ann Arbor, Michigan

Section Editor: Lalena M. Yarris, MD, MCR
Submission history: Submitted May 4, 2015; Revision received September 1, 2015; Accepted September 9, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27357

Introduction: The primary objective of this study was to determine the prevalence of remediation, competency domains for remediation, the length, and success rates of remediation in emergency medicine (EM).

Methods: We developed the survey in Surveymonkey™ with attention to content and response process validity. EM program directors responded how many residents had been placed on remediation in the last three years. Details regarding the remediation were collected including indication, length and success. We reported descriptive data and estimated a multinomial logistic regression model.

Results: We obtained 126/158 responses (79.7%). Ninety percent of programs had at least one resident on remediation in the last three years. The prevalence of remediation was 4.4%. Indications for remediation ranged from difficulties with one core competency to all six competencies (mean 1.9). The most common were medical knowledge (MK) (63.1% of residents), patient care (46.6%) and professionalism (31.5%). Mean length of remediation was eight months (range 1-36 months). Successful remediation was 59.9% of remediated residents; 31.3% reported ongoing remediation. In 8.7%, remediation was deemed "unsuccessful." Training year at time of identification for remediation (post-graduate year [PGY] 1), longer time spent in remediation, and concerns with practice-based learning (PBLI) and professionalism were found to have statistically significant association with unsuccessful remediation.

Conclusion: Remediation in EM residencies is common, with the most common areas being MK and patient care. The majority of residents are successfully remediated. PGY level, length of time spent in remediation, and the remediation of the competencies of PBLI and professionalism were associated with unsuccessful remediation. [West J Emerg Med. 2015;16(6):839–844.]

INTRODUCTION

Residency training programs have the responsibility to ensure physicians develop the knowledge, skills, and attitudes required to practice medicine independently and to measure trainees' competency.[1] It is expected that individual trainees will attain Accreditation Council for Graduate Medical Education (ACGME) Milestones at different stages during their training.[2] However, some residents will need remediation with additional resources, time and effort not typical of the majority of trainees in order to meet the established standards of each specialty training program. Much work has been done to improve the understanding and assessment of the competencies; however, few studies have addressed the impact of the competencies on remediation or the process of correcting deficiencies in trainees with the goal of graduating competent attending physicians.[3]

When program directors (PDs) identify a resident who requires additional resources to achieve the minimal competency standards in one of the six ACGME domains, it is recommended that they place that resident on a remediation pathway.[4,5] These remediation plans are tailored to the specific deficiencies of each resident, with the goal that the resident will demonstrate competency in those domains prior to graduation. However, a recent study from the members of the Council of Residency Directors (CORD)–Emergency Medicine (EM) Remediation Task Force reported great variation in the definition and management of remediation among EM programs.[4] The national prevalence of remediation, domains of concern and success rates of remediation in EM are not known.

The primary objective of this study was to determine the prevalence of remediation in EM residencies. Secondary objectives included determining the indications, length, and success rates of remediation across the EM residency programs in the United States. A better understanding of remediation will help programs to recognize possible vulnerable times in residency training, or specific domains of EM practice associated with a higher likelihood of unsuccessful resident remediation.

METHODS

The study developed an anonymous electronic survey using Surveymonkey™ that was distributed via email directly to all 160 allopathic EM PDs in the spring of 2014 (Appendix 1). We excluded two programs that indicated they were new and did not yet have any residents. Three reminder e-mails were sent to non-responders. The institutional review board reviewed this study and determined it to be exempt.

Survey Development

To provide content validity evidence, four PDs with more than 25 combined years of experience collaborated to construct the survey. The authors are integrally involved in, and provide content expertise in, the area of remediation practices, given their roles on the CORD Remediation Task Force and long-term experience as PDs and medical education leaders. Further, we formulated survey questions through a joint effort with members of the Remediation Task Force. For response process validity, questions were field tested with educational leadership faculty at the authors' programs, feedback was gathered, and questions were revised as needed.

The instrument, with specific instructions to include the last three years of data, collected the following information: program demographics; number of residents; number of residents placed on remediation (formal or informal) in the last three years; the post-graduate year (PGY) level of the resident(s) placed on remediation, length of remediation, whether or not the remediation was successful; and the core competency for which the resident was remediated.

The primary objective of this study was to determine the prevalence of remediation in EM residencies. In addition, we looked at the outcome measure of successful remediation of individual residents. Independent variables included program type (PGY-3 vs. PGY-4), training year the resident was placed on remediation, individual core competencies cited as deficient, length of time spent on remediation, and a stratified number of deficient competencies identified. The training year identified combined PGY-3 and PGY-4 into a single "senior resident" category, due to small numbers.

Outcomes and Data Analysis

Descriptive data were reported. Survey results included program size and total number of residents, which we calculated based on average class size over a three year period in order to obtain the number of residents who were at risk of remediation in the sample. Residents included in the analysis were all individuals with reported outcome data. The results were explored on the basis of inciting factors to place a resident on remediation and also factors associated with successful and unsuccessful remediation. We performed statistical analysis using STATA 12. A multinomial logistic regression model was estimated and presented in Table 1. Covariates included program length, training year resident identified, length of time on remediation, and each of the individual core competencies as identified issues, and grouping of number of identified concerns. We performed model characteristics of area under the ROC curve and Hosmer-Lemeshow test.

RESULTS

We obtained responses from 126 programs (79.7%). The majority (71%) were three-year programs, while 29% were four-year programs. Six programs were in existence for less than three years. The number of residents per program ranged from six to 84 with a total of 4,711 over the three-year period.

Remediation Prevalence and Practice

There were a total of 351 residents on remediation in

Table 1. Successful remediation compared to failed remediation. Base category is failure. Area under the Receiver Operating Characteristic (ROC) curve for success vs failure is 0.82, indicating good discriminatory power in the model. Area under the ROC curve for success vs. in progress is 0.44. The Hosmer-Lemeshow test for goodness of fit had a p<0.62, indicating non-statistically significant differences between deciles and therefore an adequate fit to the data.

Success versus failure	Relative risk ratio (standard errors)	95% CI
PGY 3 vs PGY 4 programs	1.16 (0.71)	0.35–3.83
Year identified for remediation		
PGY 2 vs PGY 1	5.15 (3.07)**	1.60–16.56
PGY 3 & 4 vs PGY1	3.29 (2.16)	0.91–11.92
Length of time in remediation in months	0.91 (0.03)*	0.85-0.98
Competency domain		
Patient care	0.04 (0.07)	0.00–1.06
Medical knowledge	0.14 (0.23)	0.01–3.41
Communication skills	0.21 (0.33)	0.01–4.51
Practice based learning	0.03 (0.06)*	0.00- -0.96
System based practice	0.20 (0.29)	0.01–3.37
Professionalism	0.03 (0.05)*	0.00–0.66
Number of identified concerns		
Two vs. one competency	25.87 (43.94)	0.93–721.97
Three vs. one competency	115.4 (357.5)	0.27–50,043.33
Four or more vs. one competency	837.7 (4,261)	0.04–17,900,000

PGY, post-graduate year
*p<0.05.
**p<0.01.

the last three years. Most programs (90%, 113) had at least one resident on remediation during the past three years, while 66% had more than one resident on remediation. The calculated prevalence of remediation in all programs was 4.4%. Remediation periods ranged from one month, while others were greater than three years (in four-year programs). The mean length of successful remediation was 8.0 (SD 5.1) months; for unsuccessful remediation it was 9.9 (SD 8.3) months, and for residents still in progress it was 8.5 (SD 5.3) months. The overall mean of time on remediation in the data was 8.2 (SD 5.5) months.

Domain of Difficulty and Year of Identification

We found that almost half of residents were identified for remediation (47.9%) during the PGY-2 year, while 26.2% were identified during the PGY-1 year. Respective characteristics of residents placed on remediation by individual competency, training year identification, and number of core competencies cited are provided in Table 2. Of the residents remediated, the three most commonly cited competencies as a concern were patient care (n=155 out of 333; 46.6%), medical knowledge (MK) (n=210 out of 333; 63.1%), and professionalism (n=105 out of 333; 31.5%). Less common competencies reported were communication skills (n=84 out of 333; 25.2%), PBLI (n=40 out of 333; 12.0%), and system-based practice (n=34 out of 333; 10.2%). One to two deficient competencies were most common (72.9%) for residents in remediation (Table 2).

PDs were asked to give specific reasons why residents were placed into remediation/probation status. Here, many individualized specific reasons were cited for changing a resident's status. However, two comments seemed to recur: performing poorly on the in-training exam (ITE), and "personality flaws," although many PDs did also comment on the fact that most of those types of issues are not really changeable.

Successful and Failed Remediation

Successful remediation was common (59.9%) and failure uncommon (8.7%), with many residents' remediation still in progress (31.3%) and thus the outcome is unknown. The multinomial logistic regression using successful remediation, failure of remediation, and ongoing remediation as the outcomes, and independent variables of program length, training year identified, length of time on remediation, patient care, MK, communication skills, PBLI, system-based practice, professionalism, and number of competencies, resulted in a statistically significant model (p<0.005).

PBLI and professionalism problems were correlated with a decreased likelihood of successful remediation. The training year at time of identification for remediation was found to be statistically significant, with later identification in residency associated with an increased relative chance for success (Table 1). This effect was most clearly demonstrated in PGY-2 vs PGY-1, with residents identified in PGY-1 having a decreased likelihood of successful remediation. Increased length of time spent in remediation was also associated with a decreased likelihood of successful remediation. There was an inverse correlation between year identified and number of competencies identified, meaning PGY-1 had fewer concerning competency domains but it had a more powerful correlation with the outcome of unsuccessful remediation compared to number of competency domains. This resulted in year of identification being significant but not number of competencies. The in-progress outcome was omitted for clarity from Table 1 as it provided no additional statistically significant findings.

DISCUSSION

Our study found that it is common for EM residencies to place residents on remediation, with 90% of programs

Table 2. Remediation characteristics. Total residents includes all residents with reported outcome data taking into account missing data.

Core competencies*	Number of residents on remediation with this issue	Successful remediation (%)	Unsuccessful remediation (%)	Still in progress (%)
Patient care	155	82 (53.3%)	18 (11.7%)	54 (35.1%)
Medical knowledge	210	127 (61.4%)	16 (7.7%)	64 (30.9%)
Communication skills	84	43 (51.2%)	8 (9.5%)	33 (39.3%)
Practice based learning	40	14 (35.0%)	8 (20.0%)	18 (45.0%)
System based practice	34	16 (47.1%)	6 (17.7%)	12 (35.3%)
Professionalism	105	51 (49.0%)	13 (12.5%)	40 (38.5%)
Issue in 1 competency	149 (45.4%)	102 (67.5%)	11 (7.3%)	36 (23.8%)
Issue in 2 competencies	105 (32.0%)	64 (61.0%)	6 (5.7%)	35 (33.3%)
Issue in 3 competencies	44 (13.3%)	20 (44.4%)	3 (6.7 %)	21 (46.7%)
Issue in 4 or more competencies	30 (9.1%)	10 (33.3%)	8 (26.7%)	12 (40.0%)
PGY 1 remediation outcome	88 (26.8%)	40 (45.5%)	11 (12.5%)	37 (42.0%)
PGY 2 remediation outcome	160 (48.8%)	103 (63.9%)	11 (6.8%)	46 (28.6%)
PGY 3 & 4	80 (24.4%)	54 (67.5%)	5 (6.3%)	21 (26.3%)

PGY, post-graduate year
*Number >100% as some residents have more than one competency identified.

reporting at least one resident on remediation in the last three years. More impressively, the data show approximately 4.4% of all EM residents on remediation during the three-year time period with 8% of these residents eventually failing the remediation process. Controlling for other variables, the year of starting remediation (intern year), increased length of remediation, and remediation in the domains of PBLI and professionalism were statistically more likely to have an unsuccessful remediation.

It is common for trainees to be on remediation for deficits in more than one competency domain. This is similar to other studies of internal medicine and pediatric residents.[6] When looking at the reasons residents were placed into remediation status, grouped by the core competencies, MK was found to be the most common domain for remediation.[5,7]

This is likely multifactorial. It may be the easiest core competency deficiency to identify, since almost all EM programs use the ITE.[8] Further standardized testing can be used to target remediation on MK by implementing an individualized education plan for low scoring residents to improve scores.[9,10] Several studies have found this to be effective.[9,10] While MK may be the most common domain, it was also found to be the most successful core competency to remediate. This high success rate is probably due to the large number of tools available to aid in the remediation process for knowledge gaps. Question banks and board review courses specifically target these issues, so personal remediation plans do not have to be created other than identifying the issue and granting access to such tools. While these are approaches to remediation, Hauer and colleagues called for more research to develop evidence-based strategies for remediation.[11] System-based practice, PBLI, and professionalism were found to be the least common reasons for residents to be placed on remediation. It is possible that this is due to difficulty with measurement. In particular, professionalism may be reported by private communication rather than an official format such as a rotation evaluation.[12,13] On the other hand, PBLI and professionalism were the competencies least likely to be successful in remediation. However, most of the residents with these deficiencies had problems with other domains as well.

The PGY-2 year was the most common time for residents to be placed on remediation. The etiology of this may again be multifactorial. It is possible that PGY-1s were less likely to be placed on remediation because PDs understand that these residents have not yet developed many skills in the core competencies. Therefore, if problems manifest in the intern year, they were significant. Additionally, many of the intern months are spent in other departments and the ITE results return late in the year. The assessment data may therefore be suboptimal. Further, second-year residents begin to have significant responsibility within the ED, allowing deficits to manifest. However, interns placed on remediation were more likely to fail remediation compared to other years, with up to 20% of interns on remediation being reported as "unsuccessful" remediation.

Residents were found to be on remediation status for a variable length of time. Successful remediation requires time to develop and implement plans, monitor resident progress and allow the resident to demonstrate improvement. Not surprisingly, residents with longer remediation were more likely to be unsuccessful. In addition, it should also be pointed out that when residents are found to be deficient in more than one core competency, their remediation plan should also be multifaceted and should target each deficiency with a specific plan to correct each gap.

While our study found remediation to be common, our results may underestimate the frequency of resident problems. Yao reported that 20% of surveyed PDs of internal medicine

residency programs reported fear of litigation and retribution as a reason for avoidance of labeling problem residents as "on remediation."[14] In addition, there is a large amount of overlap in the reasons for residents being placed into remediation. It is possible that an individual with a single deficiency in one core competency may be overlooked if the resident is strong in other competencies. Sullivan et al. give the example of the resident who is repeatedly late for conference whose unprofessional behavior may be overlooked if they excel in MK.[12] These may underestimate the frequency of behaviors that might be considered for remediation and be considered a limitation for this study. Nonetheless, once identified, residents may have multiple areas of concern they need to work on correcting.

Future directions might prospectively identify a cohort of residents on remediation and examine the overlap of domains, determine methods of successful remediation and risk factors associated with failure to remediate.

LIMITATIONS

This study had several limitations. First, it was a survey-based study with inherent limitations related to interpretation of the questions. We attempted to address the validity issues by building content and response process validity through development and piloting. Since the definitions of remediation and successful remediation are not precisely defined, there may be some variability in responses to these questions. Secondly, the total number of residents possible in the three-year period was calculated based on program size reports. This does not fully account for residents entering or leaving a program during the period and the fact that some residents will not have completed the program, but it still provides a good approximation of remediation frequency.

We compared remediation for different core competencies; however, the majority of remediation plans were for more than one competency. This makes it difficult to draw definitive conclusions about individual core competencies. Additionally, a number of the residents were currently in remediation, and the outcome for these residents is unknown. Finally, it may be difficult to remember exactly which year and what the issues were with the various residents on remediation. It was for this reason that we chose a three-year time frame, but there may be inaccuracy in response.

CONCLUSION

Resident remediation during EM residency training is common, with close to 90% of programs having at least one resident on remediation in a recent three-year period. The most common areas to remediate are MK and patient care. There is a wide range in length and success of remediation.

Address for Correspondence: John C. Burkhardt, MD, MA, University of Michigan Medical School, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-5305. Email: jburkhar@umich.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. ACGME common program requirements. Available at: https://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs2013.pdf. Accessed Oct 25, 2014.
2. Beeson MS, Carter WA, Christopher TA, et al. Emergency medicine milestones. *J Grad Med Educ.* 2013;5(1s1):5-13.
3. Wu JS, Siewert B, Boiselle PM. Resident evaluation and remediation: A comprehensive approach. *J Grad Med Educ.* 2010;2(2):242-5.
4. Weizberg M, Smith, JL, Murano T, et al.. What does remediation and probation status mean?: A survey of EM program directors. *Acad Emerg Med.* 2015;22(1):113-6.
5. Riebschleger MP and Haftel HM. Remediation in the context of the competencies: A survey of pediatrics residency program directors. *J Grad Med Educ.* 2013;5(1):60-3.
6. Dupras DM, Edson RS, Halvorsen AJ, et al. "Problem residents": Prevalence, problems and remediation in the era of core competencies. *Am J Med.* 2012;125(4):421-5.
7. Zbieranowski IM, Takahashi SG, Verma SL, et al. Remediation of residents in difficulty: A retrospective 10-year review of the experience of a postgraduate board of examiners. *Acad Med.* 2013;88(1):111-6.
8. Thundiyil JG, Modica RF, Silvestri S, et al. Do United States medical licensing examination (USMLE) scores predict in-training test performance for emergency medicine residents? *J Emerg Med.* 2010;38(1):65-9.
9. Visconti A, Gaeta T, Cabezon M, et al. Focused board intervention (FBI): A remediation program for written board preparation and the medical knowledge core competency. *J Grad Med Educ.* 2013;5(3):464-7.
10. Platt MP, Davis EM, Grundfast K, et al. Early detection of factual knowledge deficiency and remediation in otolaryngology residency education. *The Laryngoscope.* 2014;124(8):E309-E311.
11. Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Acad Med.*

2009;84(12):1822-32.
12. Sullivan C, Murano T, Comes J, et al. Emergency medicine directors' perceptions on professionalism: A council of emergency medicine residency directors survey. *Acad Emerg Med.* 2011;18(S97-S103).
13. Adams KE, Emmons S, Romm J. How resident unprofessional behavior is identified and managed: A program director survey. *Am J Obstet Gynecol.* 2008;198(6):692.e691-692.e695.
14. Yao DC and Wright SM. National survey of internal medicine residency program directors regarding problem residents. *JAMA.* 2000;284(9):1099-104.

Results from the First Year of Implementation of CONSULT: Consultation with Novel Methods and Simulation for UME Longitudinal Training

Keme Carter, MD*
Andrew Golden, BS†
Shannon Martin, MD*
Sarah Donlan, MD‡
Sara Hock, MD*
Christine Babcock, MD, MSc*
Jeanne Farnan, MD, MHPE§
Vineet Arora, MD, MAPP¶

*University of Chicago, Section of Emergency Medicine, Chicago, Illinois
†University of Chicago Pritzker School of Medicine, Chicago, Illinois
‡NorthShore University HealthSystem, Division of Emergency Medicine, Evanston, Illinois
§University of Chicago, Section of Hospital Medicine, Chicago, Illinois
¶University of Chicago, Section of General Internal Medicine, Chicago, Illinois

Section Editor: David Wald, DO
Submission history: Submitted January 23, 2015; Revision received August 30, 2015; Accepted September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.25520

Introduction: An important area of communication in healthcare is the consultation. Existing literature suggests that formal training in consultation communication is lacking. We aimed to conduct a targeted needs assessment of third-year students on their experience calling consultations, and based on these results, develop, pilot, and evaluate the effectiveness of a consultation curriculum for different learner levels that can be implemented as a longitudinal curriculum.

Methods: Baseline needs assessment data were gathered using a survey completed by third-year students at the conclusion of the clinical clerkships. The survey assessed students' knowledge of the standardized consultation, experience and comfort calling consultations, and previous instruction received on consultation communication. Implementation of the consultation curriculum began the following academic year. Second-year students were introduced to Kessler's 5 Cs consultation model through a didactic session consisting of a lecture, viewing of "trigger" videos illustrating standardized and informal consults, followed by reflection and discussion. Curriculum effectiveness was assessed through pre- and post- curriculum surveys that assessed knowledge of and comfort with the consultation process. Fourth-year students participated in a consultation curriculum that provided instruction on the 5 Cs model and allowed for continued practice of consultation skills through simulation during the Emergency Medicine clerkship. Proficiency in consult communication in this cohort was assessed using two assessment tools, the Global Rating Scale and the 5 Cs Checklist.

Results: The targeted needs assessment of third-year students indicated that 93% of students have called a consultation during their clerkships, but only 24% received feedback. Post-curriculum, second-year students identified more components of the 5 Cs model (4.04 vs. 4.81, $p<0.001$) and reported greater comfort with the consultation process (0% vs. 69%, $p<0.001$). Post- curriculum, fourth-year students scored higher in all criteria measuring consultation effectiveness ($p<0.001$ for all) and included more necessary items in simulated consultations (62% vs. 77%, $p<0.001$).

Conclusion: While third-year medical students reported calling consultations, few felt comfortable and formal training was lacking. A curriculum in consult communication for different levels of learners can improve knowledge and comfort prior to clinical clerkships and improve consultation skills prior to residency training. [West J Emerg Med. 2015;16(6):845–850.]

INTRODUCTION

Medical errors in the inpatient setting are frequently attributed to breakdowns in communication; as many as 70% of errors are attributed to communication errors.[1-6] One type of communication that is especially common in emergency department (ED) care is the consultation, whereby one provider seeks formal recommendations from another provider regarding the care of a patient; 40% of all ED visits require at least one consultation by Emergency Medicine (EM) providers.[4,7] There is increasing recognition that a formal approach to requesting consultations is necessary to prevent communication errors from occurring.[8] Curbside consultations, or unstructured consultations, whereby a consultant is asked to provide recommendations regarding the care of a patient without formal assessment and communication, have historically been a common practice in medicine.[8] However, when compared to formal consultations, curbside consultations can adversely affect patient care.[8]

There is a large gap of knowledge on effective consultation education amongst trainees and practicing physicians. Previous research focused on providing an "educational protocol" for medical students to utilize while working in the ED, including pre-developed scripts and checklists when requesting specialty consultation.[9] However, many students and trainees still receive little to no formal education specifically on consultation. A significant body of research on consultation has been developed within the specialty of EM. A conceptual model developed by Kessler et al, the "5 Cs of Consultation", has been proposed to describe a standardized consultation from the ED to hospital-based services.[10,11] The 5 Cs include Contact, Communicate, Core Question, Collaboration, and Closing the Loop, and offers specific action items for each component (Appendix A). This model has been tested and validated in a randomized controlled trial amongst EM residents and increased effectiveness of consult communication in this setting.[12,13] To our knowledge, the 5 Cs model has not been implemented into a formal undergraduate medical educational curriculum on consult communication.[4] Because the practice of learning consultation communication through single point repetition may not result in improvement of this skill, experts in the field believe that formal training in consultation communication should exist at various levels of training, including undergraduate medical education.[4,13]

The specific aims of this study were to develop, pilot, and evaluate the effectiveness of a consultation communication curriculum based on Kessler's 5 Cs model for different learner levels in undergraduate medical education that can be implemented as a longitudinal curriculum. First, we aimed to conduct a targeted needs assessment among third-year medical students on their experiences, comfort level, and instruction in calling consultations. Then, based on these results, we aimed to introduce and evaluate a curriculum in calling consultations during one academic year for second-year and fourth-year medical students. Second-year students received instruction on standardized consultation communication, and fourth-year students received a didactic and simulation based curriculum that enabled structured practice of consultation skills (Appendix B). We hypothesized that this curriculum would improve consultation communication knowledge, skills, and attitudes of the target learner groups.

METHODS

Targeted Needs Assessment

To determine the need for a formal consultation curriculum, an eight-question, paper-based, anonymous targeted needs assessment was developed internally by the authors using literature review and expert opinion. The survey was administered to third-year students at the conclusion of their clinical rotations in the Spring of 2013. The survey asked questions in four general areas: [1] previous education on calling consultations, [2] previous exposure to calling consultations, [3] current comfort level of requesting consultations and collaborating with consultants, and [4] identification of the five components to be included in effective consultations according to Kessler's 5 Cs model. The survey included yes/no questions, such as "Have you ever been instructed on how to call a consult?", as well as questions in which participants used a five point Likert-type scale, for example, to rate their "comfort requesting a consult" from Very Comfortable to Very Uncomfortable (Appendix C).

Curriculum Design and Implementation

Second-Year Medical Students

In a lecture hall setting, as one group, second-year students participated in a 50-minute consultation communication didactic during the 2014 winter quarter of a required clinical skills course. Prior to participating, all students had completed at least 30 hours of required clinical observation experiences. The authors filmed two trigger tapes for the didactic session. The first trigger tape demonstrated a curbside consultation and illustrated patient safety concerns that can arise with informal consultations. After discussion of the first trigger tape and the behaviors that led to poor consultation communication, the students were instructed on Kessler's 5 Cs model through a didactic lecture developed for all learner groups. Students then viewed the second trigger tape that illustrated a consultation that followed the 5 Cs model, and participated in a discussion of the elements of this standardized consultation that led to a successful collaborative relationship between the provider and consultant.

Fourth-Year Medical Students

Fourth-year students participating in the required, month-long EM clerkship during the 2013-2014 academic year were consented for participation. Oral consent was obtained from July 2013 through April 2014. Up to twelve students per month participate in the EM clerkship, which includes 15 hours in dedicated didactic, simulation, and educational time.

The curriculum was comprised of three parts: the didactic portion, consultation communication practice during high-fidelity simulations, and structured debriefing. A 30-minute didactic lecture was given by the EM Clerkship Director, emphasizing the importance of consultation communication and instructing students on the 5 Cs model. Following the lecture, students were provided with a pocket card detailing Kessler's 5 Cs model as an added tool for reference during their simulation sessions and clinical work, including detailed information to be included in each section of a consultation (Figure 1). During the simulation sessions, students called consultations that were recorded and reviewed by EM attending physicians who were trained using the 5 Cs Model. High-fidelity simulation was chosen as a teaching method to give students a realistic and engaging experience, as prior work has shown that although high-fidelity simulation can trigger a "stressful" response, EM trainees continue to desire to participate in future sessions.[14,15] The students then received structured feedback on their consultation communication performance during the debriefing period. Cases used in this study were selected on the basis of common presentations to the ED and included gastrointestinal bleed, myocardial infarction, ectopic pregnancy, urosepsis, diabetic ketoacidosis, aortic dissection, status epilepticus, hyperkalemia, and symptomatic bradycardia.

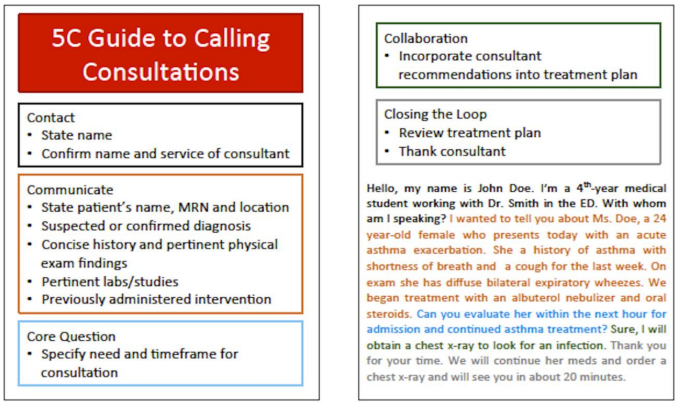


Figure 1. 5 Cs pocket card given to students participating in the consultation curriculum.
MRN, medical record number

Curriculum Evaluation

Second-Year Medical Students

Second-year students participating in the consultation communication didactic completed a pre- and post-curriculum survey. Both surveys assessed the students' knowledge of the consultation process with yes/no statements such as "I understand what a formal consultation is" as well as a multiple choice question asking them to "select the 5 components of a consultation that have been shown to improve consultation communication" according to Kessler's 5 Cs model. Both the pre- and post-curriculum surveys asked the students to "rate their level of comfort requesting a consultation given the necessary medical background." Additionally, the post-curriculum survey assessed overall satisfaction with the consultation didactic.

Fourth-Year Medical Students

Throughout the month-long clerkship, fourth-year students participated in three simulation sessions (one session per week). The initial, or baseline, simulation experience occurred prior to the implementation of the didactic component of the curriculum, and the students underwent pre- and post-curriculum evaluations. Consultation communication skills during simulation were measured by EM attending physicians who were trained using Kessler's 5 Cs Model Checklist for Assessing Physician Consultations and the Global Rating Scale (GRS) for Assessing Physician Consultations (Appendix A and D). Kessler's 5 Cs Model Checklist was adapted from a business consultation model by an expert panel and validated in a cohort of EM and EM/internal medicine residents.[12] The checklist included 13 different components that should be included in effective consultations, such as specifying the need for a consultation. The checklist components were valued as "Done" or "Not done." Kessler's GRS tool was developed through literature review and expert panel recommendations followed by review and modification by consultants.[13] The GRS utilized a five point Likert-type scale from "Not effective" to "Extremely effective" for seven items, such as patient case presentation, to indicate perceived efficacy of the consultation by the attending physician.

Each consultation was rated by two independent evaluators per consultation using the GRS and the 5 Cs Checklist. The scores for each component of the evaluations were averaged for each consultation, creating a composite, single evaluation per consultation performed.

Learner satisfaction with the fourth-year curriculum was assessed through use of a ten-question survey completed by students at the end of the EM clerkship. Learners were asked to rate the value of each curriculum component including the didactic lecture, simulated cases, and pocket cards. All survey questions were rated on a five point Likert-type scale from "Strongly disagree" to "Strongly agree."

Data Analysis

Data from the fourth-year curriculum was collected from July 2013 to April 2014. For each consultation, the 5 Cs checklist was recorded with the completion of individual checklist items, as well as having a proportion of the 12 points completed (i.e. 7/12). Responses gathered from the needs assessment, the GRS, and the learner satisfaction survey were translated into ordinal numbers for data analysis (i.e. 1="Not effective", 5="Extremely effective).

Evaluations were compared for the pre- and post-curriculum simulations. For the checklist, the proportion of inclusion of each component and the absolute difference between values before and after curriculum implementation

were reported. For the GRS, the averages and standard error of the means were reported, as well as the absolute differences in these values before and after curriculum implementation. Changes in the average checklist completion and GRS values pre- and post-curriculum implementation were analyzed using two-tailed, unpaired Student's t-tests to examine if the curriculum was overall successful at increasing thoroughness and efficacy of consultations called. Rater agreement was analyzed by calculating the distribution of differences between the two evaluators.

All statistical analyses were performed using Stata 12.1 (StataCorp, College Station, TX). A p-value of less than 0.05 was considered statistically significant in all analyses. This study was approved by the appropriate Institutional Review Boards.

RESULTS

Targeted Needs Assessment

A total of 57 third-year students out of an eligible 96 completed the targeted needs assessment, resulting in a 59% response rate. As shown in Figure 2, 53 (93%) of third-year students completing the survey reported calling a consultation during any of their third-year rotations. Thirteen students (24%) reported receiving feedback on their ability to call consultations from their supervising resident or the recipient of the consult. Forty-one students (72%) reported receiving instruction on how to call a consult, and of those, almost all (40/41, 98%) described informal instruction by a resident during their third-year rotations.

Although most students reported calling consultations, less than half (26 students, 46%) were comfortable requesting a consultation. A higher fraction felt comfortable telling the patient's story to a consultant (33 students, 58%) and receiving recommendations from a consultant (35 students, 61%). On the knowledge-based portion of the survey, roughly half (31 students, 54%) were able to correctly identify at least 4 of the 5 Cs of Kessler's model. We used this data to demonstrate the need for inclusion of formal consultation communication training in the undergraduate medical education curriculum.

Second-Year Medical Student Curriculum

Twenty-five second-year students completed the pre-curriculum survey, and 26 completed the post-curriculum evaluation. After receiving the curriculum, students were able to identify significantly more correct components of the 5 Cs model (4.04 vs. 4.81, $p<0.001$). The number of students understanding the definitions of formal and curbside consultations also significantly increased (24% vs. 100%, $p<0.001$; 16% vs. 100%, $p<0.001$; respectively). After the curriculum, more students indicated they would be comfortable in requesting consults (0% vs. 69%, $p<0.001$), while fewer felt they would need guidance while calling a consultation (60% vs. 31%, $p<0.001$). Learner satisfaction was very high, with 100% of students reporting that they were "Satisfied" or "Extremely satisfied" with the consultation didactic session.

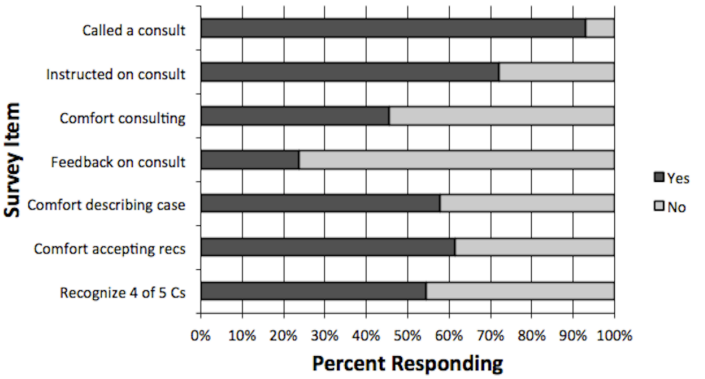


Figure 2. Percentage of students answering "yes" and "no" to each survey item of the targeted needs assessment.

Fourth-Year Medical Student Curriculum

In the fourth-year curriculum, 117 students called 170 total simulated consultations–84 prior to receiving the curriculum and 86 after curriculum implementation. Each consultation was evaluated using the GRS and the 5 Cs checklist by two independent raters per consultation for a total of 340 evaluations. Analysis of the GRS evaluations showed that in each category, evaluators differed by 2 or more points on the Likert-type scale in less than 29% of consult evaluations and were in complete agreement or differed by 1 point on the scale in each category in 72-90% of evaluations. Analysis of the 5 Cs checklist showed that evaluators differed by 2 or more criteria in each category in less than 11% of consult evaluations and were otherwise in complete agreement or differed by 1 criterion in each category in 90-100% of evaluations. The scores for each component of the evaluations were averaged for each consultation, creating 170 composite evaluations.

As shown in Figure 3, when compared to pre-curriculum consultation evaluation, the combined average score of the criteria measured in the GRS, or Average GRS, increased significantly following the implementation of the curriculum (3.05 vs. 3.70, $p<0.001$). Additionally, consultations performed after the implementation of the curriculum scored significantly higher in all seven individual criteria on the GRS ($p<0.001$ for all).

After receiving the curriculum, students completed higher proportions of the 5 Cs checklist compared to their consultation evaluations prior to receiving the curriculum (Figure 4). The overall checklist completion increased significantly (62% vs. 77%, $p<0.001$); specifically, the Contact (30% vs. 59%, $p<0.001$), Communicate (85% vs. 90%, $p<0.05$), and Closing the Loop (74% vs. 89%, $p<0.001$) sections had significantly higher completion by the students after the implementation of the curriculum.

Sixty-nine (59%) of the 117 fourth-year students who received the curriculum completed the learner satisfaction

survey. Participants reported a high degree of satisfaction with the curriculum, as 99% (68/69) indicated being "Satisfied" or "Extremely satisfied." When asked about the content of the curriculum, 100% of the students "Agreed" or "Strongly Agreed" that the information presented had not been taught previously in their medical education. About 94% (65/69) of students reported the simulation experiences helped prepare them for calling consultations in a clinical setting. Almost all (67/69, 97%) students rated the pocket card as useful. Finally, after the implementation of the curriculum, 94% (65/69) of students "Agreed" or "Strongly Agreed" with the statement, "I feel comfortable calling a consultation."

DISCUSSION

Through our curriculum, second-year students are introduced to the formal consultation process, and instructed on a standardized consult model prior to beginning clerkships. Fourth-year students are instructed on the 5 Cs model, given a pocket card that encourages continued adherence to the standardized consultation, and practice their consultation skills in a structured, simulation setting. After participating in our consultation curriculum, second-year students demonstrated an increase in knowledge and understanding of the standardized consultation process and, although not surprising given their lack of significant clinical experience, reported higher levels of comfort with requesting a consultation given the necessary medical background. Fourth-year students scored higher on evaluations that assess thoroughness and perceived efficacy of consultation communication and reported higher levels of comfort in calling and discussing consultations compared to responses gathered in the targeted needs assessment. These

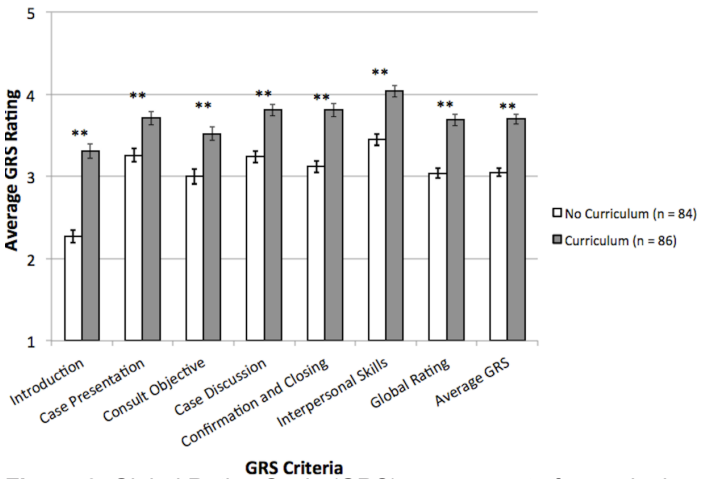


Figure 3. Global Rating Scale (GRS) assessment of consultation efficacy in seven criteria and an average of all criteria. Five point scale responses were converted to ordinal numbers where 1="Not effective" and 5="Extremely effective". Means are graphed before and after curriculum implementation with standard error of the mean error bars.
**p<0.001.

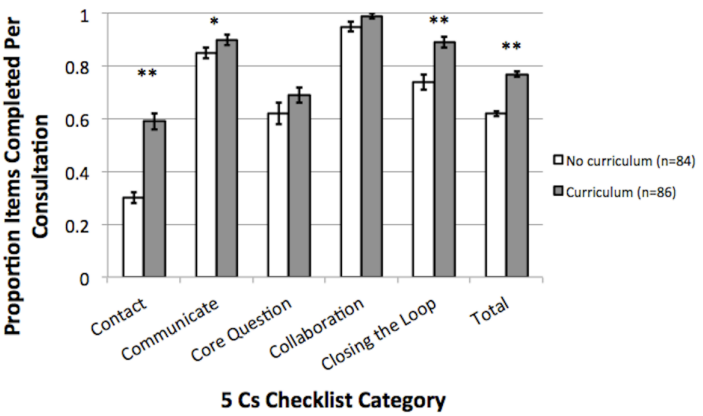


Figure 4. Proportion of 5 Cs checklist items completed per consultation in each category and total completion. For before and after the curriculum implementation, proportions are graphed with standard error of the mean error bars.
*p<0.05.
**p<0.001.

results suggest that on-the-fly instruction in consultation communication is not adequate, and a formal curriculum is needed to improve skills and comfort level.

These findings have implications for the inclusion of additional educational interventions in undergraduate medical education to enhance the efficacy of consultation communication. This consultation communication curriculum, which can be implemented as a longitudinal experience, is novel and allows students to receive instruction prior to the third-year clerkships. It also reinforces this critical skill in the controlled environment of the simulation laboratory during a fourth year capstone experience. Our curriculum aligns well with the need for medical schools to address whether students are proficient in the Core Entrustable Professional Activities for Entering Residency, particularly those emphasizing the importance of interprofessional collaboration, understanding one's role as a medical team member, and seeking help when necessary,[16] which our curriculum specifically addresses. Although the 5 Cs Model was validated in a cohort of resident physicians, we believe our consultation curriculum is feasible, and was shown to be effective and suitable for this level of learner.

Future directions in assessing the effectiveness of this consultation curriculum include evaluating long-term retention of consultation communication knowledge and skills in a group of learners who receive the entire longitudinal curriculum. We plan to study consultants' perceptions of consults called by students to determine if these skills translate to the clinical setting. Also, we plan to compare the consultation communication skills of interns who have received the longitudinal curriculum as medical students to entering interns who have not received consultation training, allowing for a control group study.

LIMITATIONS

There are limitations to this study. Our research and

educational curriculum were performed at a single site, and thus results may not be generalizable to other institutions. The response rate for the targeted needs assessment and fourth-year learner satisfaction surveys was 59%. As the surveys were anonymous, we were unable to follow up with students who did not complete the surveys to increase the response rate. Additionally, the analysis of consultations occurred only in controlled environments. In real clinical settings, the measures of an effective consultation according to Kessler's model may be sacrificed for issues of timeliness, and other responsibilities of serving on a care team. Finally, our study describes interventions at the second and fourth-year levels. During year 1 of curriculum implementation, we were unable to study the same learners over time to determine if receiving the entire curriculum has any long-term educational benefit, but plan to do so in the future.

CONCLUSIONS

Medical students are calling consultations during third-year clerkships, but formal instruction is rare. Developing and implementing longitudinal consultation curricula, with a didactic during the pre-clinical curriculum and simulation-based instruction during the EM clerkship, can help address the current deficit in undergraduate medical education and better prepare students to call consults before beginning clinical clerkships and prior to entering residency training.

ACKNOWLEDGMENTS

The Academy of Distinguished Medical Educators of the Pritzker School of Medicine, the faculty and staff at the Grainger Center for Simulation and Innovation at NorthShore University HealthSystem, and Kristen Wroblewski, MS.

Address for Correspondence: Keme Carter, MD, University of Chicago, 5841 South Maryland Avenue, MC 5068 Chicago, IL 60637. Email: kcarter@medicine.bsd.uchicago.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Cheung DS, Kelly JJ, Beach C, et al. Improving Handoffs in the Emergency Department. *Ann Emerg Med.* 2010;55(2):171-80.
2. Eisenberg EM, Murphy AG, Sutcliffe K, et al. Communication in Emergency Medicine: Implications for Patient Safety. *Commun Monogr*. 2005;72(4):390-413.
3. Horwitz LI, Meredith T, Schuur JD, et al. Dropping the Baton: A Qualitative Analysis of Failures During the Transition From Emergency Department to Inpatient Care. *Ann Emerg Med.* 2009;53(6):701-10.e4.
4. Kessler CS, Chan T, Loeb JM, et al. I'm Clear, You're Clear, We're All Clear: Improving Consultation Communication Skills in Undergraduate Medical Education. *Acad Med J Assoc Am Med Coll.* 2013;88(6).
5. Maughan BC, Lei L, Cydulka RK. ED handoffs: observed practices and communication errors. *Am J Emerg Med.* 2011;29(5):502-11.
6. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. *Acad Med J Assoc Am Med Coll.* 2004;79(2):186-94.
7. Woods RA, Lee R, Ospina MB, et al. Consultation outcomes in the emergency department: exploring rates and complexity. *CJEM.* 2008;10(1):25-31.
8. Burden M, Sarcone E, Keniston A, et al. Prospective comparison of curbside versus formal consultations. *J Hosp Med*. 2013;8(1):31-5.
9. Go S, Richards DM, Watson WA. Enhancing medical student consultation request skills in an academic emergency department. *J Emerg Med.* 1998;16(4):659-62.
10. Kessler CS, Asrow A, Beach C, et al. The taxonomy of emergency department consultations--results of an expert consensus panel. *Ann Emerg Med.* 2013;61(2):161-6.
11. Kessler CS, Kutka BM, Badillo C. Consultation in the emergency department: a qualitative analysis and review. *J Emerg Med.* 2012;42(6):704-11.
12. Kessler CS, Kalapurayil PS, Yudkowsky R, et al. Validity evidence for a new checklist evaluating consultations, the 5Cs model. *Acad Med J Assoc Am Med Coll.* 2012;87(10):1408-12.
13. Kessler CS, Afshar Y, Sardar G, et al. A prospective, randomized, controlled study demonstrating a novel, effective model of transfer of care between physicians: the 5 Cs of consultation. *Acad Emerg Med Off J Soc Acad Emerg Med.* 2012;19(8):968-74.
14. Gordon JA. As accessible as a book on a library shelf: the imperative of routine simulation in modern health care. *Chest*. 2012;141(1):12-6.
15. Kharasch M, Aitchison P, Pettineo C, et al. Physiological stress responses of emergency medicine residents during an immersive medical simulation scenario. *Dis--Mon DM*. 2011;57(11):700-5.
16. Association of American Medical Colleges. Core Entrustable Professional Activities for Entering Residency, Curriculum Developer's Guide. 2014.

Does the Concept of the "Flipped Classroom" Extend to the Emergency Medicine Clinical Clerkship?

Corey Heitz, MD*
Melanie Prusakowski, MD*
George Willis, MD†
Christopher Franck, PhD‡

*Virginia Tech Carilion School of Medicine, Department of Emergency Medicine, Roanoke, Virginia
†University of Maryland College of Medicine, Department of Emergency Medicine, Baltimore, Maryland
‡Virginia Polytechnic Institute and State University, Department of Statistics, Blacksburg, Virginia

Section Editor: Jonathan Fisher, MD
Submission history: Submitted April 30, 2015; Revision received August 26, 2015; September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27256

Introduction: Linking educational objectives and clinical learning during clerkships can be difficult. Clinical shifts during emergency medicine (EM) clerkships provide a wide variety of experiences, some of which may not be relevant to recommended educational objectives. Students can be directed to standardize their clinical experiences, and this improves performance on examinations. We hypothesized that applying a "flipped classroom" model to the clinical clerkship would improve performance on multiple-choice testing when compared to standard learning.

Methods: Students at two institutions were randomized to complete two of four selected EM clerkship topics in a "flipped fashion," and two others in a standard fashion. For flipped topics, students were directed to complete chief complaint-based asynchronous modules prior to a shift, during which they were directed to focus on the chief complaint. For the other two topics, modules were to be performed at the students' discretion, and shifts would not have a theme. At the end of the four-week clerkship, a 40-question multiple-choice examination was administered with 10 questions per topic. We compared performance on flipped topics with those performed in standard fashion. Students were surveyed on perceived effectiveness, ability to follow the protocol, and willingness of preceptors to allow a chief-complaint focus.

Results: Sixty-nine students participated; examination scores for 56 were available for analysis. For the primary outcome, no difference was seen between the flipped method and standard (p=0.494.) A mixed model approach showed no effect of flipped status, protocol adherence, or site of rotation on the primary outcome of exam scores. Students rated the concept of the flipped clerkship highly (3.48/5). Almost one third (31.1%) of students stated that they were unable to adhere to the protocol.

Conclusion: Preparation for a clinical shift with pre-assigned, web-based learning modules followed by an attempt at chief-complaint-focused learning during a shift did not result in improvements in performance on a multiple-choice assessment of knowledge; however, one third of participants did not adhere strictly to the protocol. Future investigations should ensure performance of pre-assigned learning as well as clinical experiences, and consider alternate measures of knowledge. [West J Emerg Med. 2015;16(6):851–855.]

INTRODUCTION

Emergency medicine (EM) provides students with the opportunity to care for undifferentiated patients, but the unscheduled and acute nature of the specialty makes it

difficult to standardize student experiences.[1] The variety of learning styles and differences in both medical knowledge and level of motivation among medical students in a mandatory clerkship are further complicated by varying patient chief complaints, levels of patient acuity, opportunities for procedures, and attending management styles. Thus, linking educational objectives and clinical learning during clerkships can be difficult.

The Clerkship Directors in EM (CDEM) created a set of recommended objectives and curricular goals for a required fourth-year EM clerkship in an attempt to standardize EM clerkships nationally.[2,3] Educators in EM historically have created uniform didactics in an attempt to standardize the medical knowledge imparted in clerkships.[4] With the understanding that there is a body of knowledge needing to be gained despite clinical variability and the unscheduled nature of emergency patient visits, some educators have employed asynchronous learning activities to which the students have access between shifts.[5,6] This concept essentially employs the objectives of pre-learning[7] and the exposure to standardized learning materials, such as computer-based learning modules, pre-recorded didactic lectures, selected literature, and preferred free open-access medical education materials.[6]

Adult learning theory places heavy emphasis on the applicability of gained knowledge.[8] The concept of "inverted" or "flipped" education was developed to enhance education application. The "flipped classroom" relies on technology or other methods of information dissemination to introduce students to course content outside of the classroom so they can employ, apply or engage that information more deeply inside the classroom.[9]

During clinical clerkships, the opportunity to apply medical knowledge occurs during time in the wards, clinic, operating room, or emergency department (ED).[3] One of the overarching objectives of the EM clerkship is to provide students with the ability to manage the undifferentiated patient.[2,3] Previous authors have demonstrated that exposure to a favorable patient mix coincides with increased confidence in managing the undifferentiated patient.[10] The challenge remains providing EM students with at least a minimum standardized exposure to high yield chief complaints that are expected to be encountered during an EM rotation and are tested in the standardized clerkship written examination.[11] Prior works show that students can be directed to standardize their clinical experiences[12] and that limiting the scope from one that is unfocused and unpredictable to one that creates areas of concentration makes it easier to focus on specified learning objectives, and this improves performance on examinations.[13]

We attempted to combine the experience of standardized, technology-assisted pre-learning with the previously-described quasi-standardized clinical experience to create a "flipped clerkship." The objectives were to create an educational method in which students were directed to learn a topic prior to an assigned shift and then focus on patients with that chief complaint during their shift. We hypothesized that this unique model for directing both asynchronous learning and the clinical experiences in the ED would result in improved medical knowledge as measured by a series of multiple-choice questions.

METHODS

Study Setting and Participants

This was a multicenter study conducted at two academic sites, Virginia Tech Carilion School of Medicine and the University of Maryland School of Medicine, between July 1, 2013 and June 30, 2014. The study participants were either late third-year or fourth-year medical students enrolled in the required EM rotation or the EM elective at either site. All participants underwent informed consent at the beginning of the rotation, which included the assurance that participation in the study was inconsequential to their final grade on the rotation.

Study Protocol

Study participants were randomly designated a study number that assigned them to a combination of two chief complaints commonly seen in the ED setting. The chief complaints were chest pain (CP), abdominal pain, (AP) dyspnea (SOB), and altered mental status (AMS). Once the participants were assigned to one of the six combinations of chief complaints (e.g., CP+SOB), they were told to choose two shifts during their rotation that would be "themed" shifts.

During themed shifts, participants would focus their attention on evaluating and managing patients who presented to the ED with the assigned chief complaint for that themed shift, with a goal of evaluating at least three patients with that chief complaint. They were instructed to select one shift for each of their two chief complaints to be the themed shift. Prior to the themed shift, participants were instructed to complete an interactive computer-based learning module discussing the chief complaint assigned for the themed shift. These web-based modules taught the subject material for how to evaluate patients with each of the chief complaints. These modules consisted of lecture material, web-based reading material, and questions specific to the chief complaint and were based heavily on the curriculum recommended by the CDEM and found at www.cdemcurriculum.org.[2] All of the participants were given access to all of the learning modules at the beginning and were permitted to use them for learning purposes at any point during their rotation. The participants were instructed to perform the learning modules for the other two chief complaints (which they were not assigned to be "themed") at a time of their choosing during the rotation.

At the end of the four-week rotation, the participants were asked to complete a 40-question examination (Appendix 1). The examination contained 10 peer-reviewed multiple-choice questions for each of the four chief complaints; students were required to answer all 40 questions (20 from their assigned themed topics, 20 from standard). Participants' performance

on the examination was not considered toward their final grade on the rotation. The examination was administered through www.saemtests.org using Logic eXtension Resourses 6.0 (LXR; Applied Measurement Professionals, Inc., Georgetown, SC) to simplify administration and tallying of results. All participants took the same examination, although the order of questions was altered by the testing software to minimize chances of unethical behavior.

After the examination, participants were asked to complete a survey. The survey asked participants to evaluate several aspects of the flipped classroom technique on a five-point scale. Points of evaluation included comparisons between the flipped clerkship and traditional learning modalities, ability to focus on the chief complaint during themed shifts, willingness of faculty and residents to allow participants to adhere to protocol, and how closely participants stuck to the protocol.

Data Analysis

Our primary analysis was to compare scores for a flipped clerkship versus standard learning. To determine whether any observable score differences were attributable to location, site (VTC/UMD) was modeled along other predictors in a mixed model framework for both the primary (flip versus standard) and secondary (topic comparisons) analyses. The mixed model approach considers observed scores as a function of flipped status, topic, and location while modeling variability among individual students as a random effect. We removed site from analyses in which it did not exhibit a statistical association with score after accounting for other model terms.

For the primary analysis, site did not show a significant association, and therefore the comparison between flipped and standard scores was accomplished using a paired t test, which assessed flipped status in a manner equivalent to the mixed model by differencing scores between flipped and standard topics for each student. We computed group means, the t statistic, p-value, and a confidence interval for the primary analysis. The distribution of the differences was assessed graphically for normality to ensure the appropriateness of this test.

For the secondary analysis, we used the mixed model approach to compare scores among topics CP, SOB, AB, and AMS. Normality of the residuals was confirmed graphically. Topic means, p-values comparing topics, and 95% confidence intervals were computed. To determine whether the flipped approach benefitted certain topics more than others, we included a statistical interaction effect between topic and flipped status in the mixed model. Observed p-values below α=0.05 are described as statistically significant in this report.

Question performance is reported as pdiff (a measure of the difficulty of the question, with 1 signifying 100% of students answering correctly) and point biserial correlation (rpb, a measurement of the ability of a question to discriminate between high overall examinees and lower overall examinees. A higher rpb is ideal and a negative rpb signifies a flawed question.)

This study was approved by the Carilion Clinic and University of Maryland Institutional Review Boards.

RESULTS

Sixty-nine students participated in the protocol. Examination scores were missing for 12 students who did not take the examination, and one data set had a missing participant identification number. Data for 56 students were included in the analysis. Twenty-one participants were fourth-year medical students on an elective EM rotation; 35 were students on a required EM rotation. Of these 35, 11 were late third-year students. Rotation length was four weeks for all students.

Overall Flipped vs. Standard Score Comparison

Each student answered 20 questions on topics they were assigned to flip, and 20 questions on topics that they prepared for in the standard fashion. Site of rotation exhibited no association with scores (p=0.3861), so the paired t test was used. We saw no statistical difference when comparing scores on flipped topics vs standard topics. The mean flipped score was 14.14, and the mean standard score was 13.89 (t=-0.69 on 55 df, p=0.494, 95% CI of difference: -0.98 to 0.48). When performing the primary analysis (overall flipped vs standard score) at each participating institution individually, no difference was found (VTC: 36 students, p=0.8959; UMD: 20 students, p=0.3927). When including data from only the 28 students who replied that they followed the protocol, there was not a statistical difference between flipped and standard scores (p=0.8071). We saw no statistical difference when comparing students on required compared to elective rotations.

Topic Comparisons

Site did not have a statistical association with score in the mixed model (p=0.3835). Statistical differences were observed between scores on the four topics. Noting that there were 10 points available per topic, the mean topic scores were AP: 7.14, AMS: 6.77, CP: 7.68, and SOB: 6.45. Scores on CP were statistically higher than the other three topics (CP:SOB, p<0.0001, 95% CI [0.76 to 1.70]; CP:AB, p=0.0251, 95% CI [-1.00 to -0.07]; CP:AMS, p=0.0002, 95% CI [-1.38 to -0.44]), and scores on AB were higher than on SOB but not AMS (AB:SOB, p=0.0038, 95% CI [0.23 to 1.16]; AB:AMS, p=0.1154, 95% CI [-0.09 to 0.84]). Scores on AMS and SOB did not significantly differ (p=0.1768, 95% CI [-0.15 to 0.79]).

Flip Benefit for Certain Criteria

No topic benefited from being flipped when compared to the other topics. This was assessed using a statistical test for interaction between flipped status and topic (p=0.167).

Question Performance

Average pdiff for examination questions was 0.69. Average rpb for examination questions was 0.37.

Student Feedback

Forty-five students completed a feedback survey at the end of the rotation. Students were asked to rate the flipped method as a learning tool compared to standard; the average rating was 3.48 out of 5 with 1 being "poor/worse" and 5 being "excellent/better." In addition, they were asked to rate their ability to evaluate patients with the assigned chief complaints, i.e., their ability to focus their shifts, with a rating of 2.66 out of 4. Most (68.9%) of respondents answered that they followed the protocol, and 31.1% responded that were not able to. The most common responses to a follow-up question of "why did you not follow the protocol?" included forgetting at the time of the shift and an inability to see patients having the chief complaint on which they were supposed to focus during that shift.

DISCUSSION

Our results suggest that numerous challenges exist with asynchronous, targeted clinical learning in the EM student clerkship. Students participating in our flipped clerkship did not show improvement on learning, as measured by multiple-choice questioning, for specific EM topics performed in a targeted fashion when compared with those who did not similarly target their clinical learning.

Prior studies have shown both a benefit in medical education from asynchronous learning, as well as areas in which the methods did not result in differences when compared with traditional learning.[13] Much of this research has been performed in the didactic setting, and most studies look at short-term, pre-test and post-test performance; few studies address retention. Our study assessed medium-term knowledge recall over the course of a clinical clerkship.

Clinical learning may also not fully match what students learn from textbooks or other standardized learning material, and this could have an effect on improvements in, and measurement of, knowledge. The clinical environment is variable, and students are exposed to variations in care.[1,12] The effects of standardizing clinical exposure and assigning reading has had differing effects on knowledge-based assessment and clinical performance.[13,15] The implications of this may be that providing a targeted, "flipped classroom" style educational approach in the clinical setting may be difficult given the unclear connection between clinical learning and knowledge gains as measured by examinations.

LIMITATIONS

A large portion (>30%) of students did not fully engage in a themed shift or reported they did not complete their asynchronous pre-learning as per the protocol. We did not collect detailed data regarding which part of the protocol was violated. From comments provided, some students did not complete the pre-shift learning, others were unable to focus on the chief complaint during a shift, and some performed more "themed" shifts than were assigned, but the proportions of each are not known. It is unclear what effect this had on the outcomes, but the concept of pre-assigned asynchronous learning and its potential benefits is directly related to the expectation that learners complete the assignments prior to the learning session.

The clinical environment poses distinct challenges. In many flipped classroom settings, the instructor is able to control the in-classroom learning session, but this may not be possible in the clinical setting. While we encouraged students to focus on topically-appropriate patients in themed shifts, we did not keep track of their patients, and students ranked the challenge of focusing on themed topics as the greatest challenge on the feedback survey. It was our hope that the act of focused learning for the purposes of preparing for an upcoming shift would fulfill the adult learning principles of creating relevancy and goal-orientation, and therefore increase learning potential. It is possible that without the reinforcement of using the information learned (i.e., seeing patients with that chief complaint), that there is little to no benefit to prior preparation.

In addition, there was no attempt to regulate when the learners performed the modules, how much effort was put into them, or what sort of education took place during their shifts. There may have been significant heterogeneity in effort and timing of module completion; for instance, some students may have performed the module the night before a shift, while others performed it several days prior. Anecdotally, the modules take up to two hours to complete, but some students may have spent significantly less time and effort, thereby affecting their efficacy. While this unscheduled aspect of asynchronous learning is one of its inherent benefits, it may also contribute to inconsistent outcomes.

At least one student noted in written feedback that he felt the method to be useful, and therefore performed all of his shifts in a "themed" fashion. If other students similarly extended the use of themed shifts to unassigned chief complaints, this may have contributed to the lack of difference in outcomes.

The possibility of Type II error exists; the pooled standard deviation between the flipped and standard scores was 2.72, the intraclass correlation for this analysis was 0.38, and the difference in scores between the flipped and standard settings was 0.25 points. If this combination of means, standard deviations, and associations were the true state of the universe, then it would take n=743 study participants to declare statistical significance between the interventions. However, the question would remain as to what size difference would reflect a true knowledge difference.

Finally, the assessment used was a non-standardized, small set of multiple-choice questions. Multiple-choice questions may not be the ideal way to assess clinical learning,[16] and while attempts were made to ensure question validity by expert consensus, the average difficulty was fairly low and several questions were very difficult (pdiff <0.5).

However, all questions had good discriminatory value. Future work may include mixed methodology including qualitative methods and could compare entire rotation blocks performed in "flipped" fashion to those performed in standard fashion, allowing several measures of overall clerkship performance to be assessed.

CONCLUSION

Preparation for a clinical shift with pre-assigned, web-based learning modules followed by an attempt at chief complaint-focused learning during a shift did not result in improvements in performance on a multiple-choice assessment of knowledge; however, one third of participants did not adhere strictly to the protocol. Future investigations should ensure performance of pre-assigned learning as well as clinical experiences, and consider alternate measures of knowledge.

Address for Correspondence: Corey Heitz, MD, Virginia Tech Carilion School of Medicine, 1 Riverside, 4th Floor, Roanoke, VA 24014. Email: crheitz@carilionclinic.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Coates WC and Gill AM. The Emergency Medicine Subinternship-A Standard Experience for Medical Students? *Acad Emerg Med*. 2001;8:253-8.
2. Manthey DE, Ander DS, Gordon DC, et al. Emergency Medicine Clerkship Curriculum: An Update and Revision. *Acad Emerg Med*. 2010;17:638-43.
3. Manthey DE, Coates WC, Ander DS, et al. Report of the Task Force on National Fourth Year Medical Student Emergency Medicine Curriculum Guide. *Ann Emerg Med*. 2006;47:e1-e7.
4. De Lorenzo RA, Mayer D, Geehr EC. Analyzing Clinical Case Distributions to Improve an Emergency Medicine Clerkship. *Ann Emerg Med*. 1990;19:746-51.
5. Burnette K, Ramundo M, Stevenson M, et al. Evaluation of a Web-Based Asynchronous Pediatric Emergency Medicine Learning Tool for Residents and Medical Students. *Acad Emerg Med*. 2009;16Suppl2:S46-S50.
6. J. Wesley Baker. "The "Classroom Flip": Using Web Course Management Tools to Become the Guide by the Side" Selected Papers from the 11th International Conference on College Teaching and Learning (2000): 9-17. Available at: http://Works.Bepress.Com/J_Wesley_Baker/21.
7. Crouch CH and Mazur E. Peer Instruction: Ten Years of Experience and Results. *Am J Phys*. 2001;69(9):970-7.
8. Merriam SB. Updating Our Knowledge of Adult Learning. *J Cont Educ Hlth Prof*. 1996;16(3):136-43.
9. Strayer Jf. How Learning in an Inverted Classroom Influences Cooperation, Innovation and Task Orientation. *Learn Environ Res*. 2012;15:171-93.
10. De Jong J, Visser M, Van Dijk N, et al. A Systematic Review of the Relationship Between Patient Mix and Learning in Work-Based Clinical Settings.
11. Senecal EL, Heitz C, Beeson MS. Creation and Implementation of a National Emergency Medicine Fourth-Year Student Examination. *J Emerg Med*. 2013;45:924-34.
12. Coates WC, Gendy MS, Gill AM. Emergency Medicine Subinternship: Can We Provide a Standard Clinical Experience? *Acad Emerg Med*. 2003;10:1138-41.
13. Lampe CJ, Coates WC, Gill AM. Emergency Medicine Subinternship: Does a Standard Clinical Experience Improve Performance Outcomes? *Acad Emerg Med*. 2008;15:82-5.
14. Jordan J, Jalali A, Clarke S, et al. Asynchronous vs Didactic Education: It's Too Early to Throw in the Towel on Tradition. *BMC Med Educ*. 2013;13:105.
15. Hoffman L, Bott K, Puumala S, et al. Influence of Assigned Reading on Senior Medical Student Clinical Performance. *West J Emerg Med*. 2009;10:23-9.
16. Newble Di, Baxter A, Elmslie Rg. A Comparison of Multiple-Choice Tests and Free-Response Tests in Examinations of Clinical Competence. *Med Educ*. 1979;13:263-8.

Assessing the Impact of Video-based Training on Laceration Repair: A Comparison to the Traditional Workshop Method

Nicholas Chien, BS
Terren Trott, MD
Christopher Doty, MD
Brian Adkins, MD

University of Kentucky, Department of Emergency Medicine, Lexington, Kentucky

Section Editor: Jonathan Fisher, MD
Submission history: Submitted May 1, 2015; Revision received August 25, 2015; Accepted September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27369
[West J Emerg Med. 2015;16(6):856–858.]

INTRODUCTION

Medical school curricula in the United States have been more recently focusing on the early integration of clinical sciences and clinical experiences into medical students' pre-clinical years. For many medical students, the common mode of instruction for developing the procedural skill of laceration repair is largely from live workshop training requiring a significant amount of physical resources and physician time to train the students. This study compares the effectiveness of video-based learning (VBL) to traditional live workshop learning (LWL) on student laceration repair performance.

Review of literature

We performed a review of the literature with searches in PubMed such as video, suturing, medical education, learning, medical students and found several relevant articles published in the last 10 years. Several studies have investigated integrating video modules into medical curriculum and overall, findings have been controversial. One study aimed to identify willingness to learn from video modules in virtual patient encounters. A total of 120 students took a post-encounter survey with majority preferring text-based learning over video. However, the video modules were perceived to be more thorough and with higher detail. In a second study, third-year medical students used VBLs for their pediatrics rotation and video modules were statistically associated with higher recognition of principal symptoms, appropriate diagnosis and consistency between observed symptoms and diagnosis. No studies have been found that used video modules for suturing technique.

METHODS

We invited first-year medical students at the University of Kentucky College of Medicine to participate in the laceration repair study. Inclusion criteria included students with no prior suturing experience and who were available to attend training (August 26, 2014) and two assessments (September 2 and November 11, 2014). Students were asked to adhere to a set of study rules where they not allowed to discuss the laceration repair study with classmates, attempt to contact other members of the study, and discuss or identify learning resources with classmates. We enrolled the first 40 students to confirm their eligibility and reply via email.

Students were randomized into two groups: VBL and LWL (Figure 1). Randomization was performed by assigning students a number, between 1 and 40, based on their order of enrollment. Students were then separated into two groups defined by odd and even assigned numbers. Students with even numbers were assigned to the VBL group, while students with odd numbers were assigned to the LWL group. On the day of training, study participant were provided one banana, one scissor, four Ethicon 4-0 silk sutures, one Addison forcep, and one needle driver as tools to practice suturing techniques. Participants were asked to keep their practice materials for the remainder of the study.

We developed a suture task checklist (Figure 2) by combining various assessment criteria used in the evaluation of suturing as published in Assessing Surgical Skill Using Bench Station Models (Khan et al.) and Clinical skills training: developing objective assessment instruments (Conner, H.M. and McGraw, R.C.). Workshop content was solely based off assessment criteria from the suture task checklist.

The live workshop was presented in Microsoft PowerPoint format by a second-year EM resident and was recorded live using Echo360 lecture capture software. An adjustable camera toggled by the presenter was used to capture imaging of the instructors hands while performing

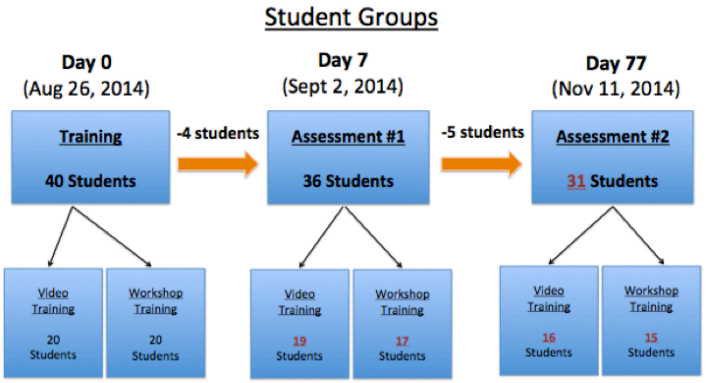


Figure 1. Student groups with randomization to video training and workshop training as on days 0, 7, 77.

suturing technique. The video recording of the workshop was posted for the 20 students in the VBL group, leaving both groups with the same instructional content.

The live workshop consisted of a 20-minute lecture followed by one hour and 40 minutes of practice and instructional feedback. There were a total of three resident physicians, including the instructor, who provided instructional feedback and tips to students during their allotted practice time. Students were not permitted to ask questions during the lecture as the lecture was being recorded for the video-training arm of the study. Students were free to ask questions during their 1hr 40min practice session and allowed to leave at any time during their practice session.

Two faculty physicians from the University of Kentucky Department of Emergency Medicine, Dr. Christopher Doty and Dr. Brian Adkins, generously volunteered their time to provide mentorship, project oversight, and assessment of student suturing performance using the 22-point suture task checklist. Both physicians were blinded to participant group assignment during the two student assessments and were present for the entirety of each assessment. The average of the two independent numerical values derived from each physician's 22-point suture task checklist was taken and used as the student's final score. We used a Welch two Sample t-test to compare performance between groups.

RESULTS

For the first assessment, 36 students were evaluated. The LWL group (n=17) scored a mean of 18.59 (SD 1.8, 95% CI [17.6-19.3]); while the VBL group (n=19) scored a mean of 18.21 (SD 1.8, 95% CI [17.3-19.0]) (p-value 0.549) (Figure 3). For the delayed assessment, 31 students were evaluated. The LWL (n=15) scored a mean of 17.87 (SD 2.5, 95% CI [16.6-19.1]); while the VBL group (n=16) score a mean of 17.75 (SD 2.5, 95% CI [16.6-19.0]) (p-value 0.8979).

Evaluator concordance using 22-point suture task checklist was as follows: Evaluators' assessment scores were identical 29.9% of the time; evaluators' assessment scores differed by one point 44.8% of the time. Therefore, evaluators scored students within one point of each other 74.7% of the time.

Suture Task Checklist	Yes	No
1. Recognizes that the wound should be sutured		
2. Enquires re: tetanus status		
3. Mentions anesthetic		
4. Sterile technique (gloves)		
5. Adequate irrigation		
6. Selection of appropriate instruments and suture		
7. Correct placement of needline in needle driver		
8. Perpendicular penetration and exit		
9. Bite no closer than 0.5cm		
10. Equal bites on either side of wound		
11. Curvature of needle followed		
12. Smooth passage of needle, no hesitancy		
13. Instrument tie technique		
14. Initial double wrap throw		
15. Square knot		
16. At least 3 knots		
17. Leaves 0.5cm after cutting suture		
18. Minimum 3 sutures		
19. Stitch perpendicular to wound edge		
20. Adequate eversion of wound edge		
21. Discuss correct time for removal (prompted by instructor) (face, hands=5-7 days. Rest=7-10 days)		
22. Wound care		

Figure 2. Suture task checklist. Suture Task Checklist derived from: Khan et al. Assessing Surgical Skill Using Bench Station Models. Conner HM and McGraw RC. Clinical skills training: developing objective assessment instruments.

DISCUSSION

Medical students often use shadowing experiences, simulation labs, and live workshops to develop procedural skills such as laceration repair that will better prepare themselves for their clinical rotations. Many of these experiences require a tremendous amount of training resources (physician time, space, practice materials, and live tissues) and planning to synchronize the availability of students and physicians. In our study, students who participated in VBL had no significant difference in suturing scores at one and three months compared to LWL. These results suggest that VBL may be as effective as live workshop training. The implementation of accessible VBL into medical students' pre-clinical education may be an effective way to teach students procedural skills while saving time, space, and resources used for scheduled instruction in an environment of ever-increasing educational demands.

While VBL serves as a promising educational tool,

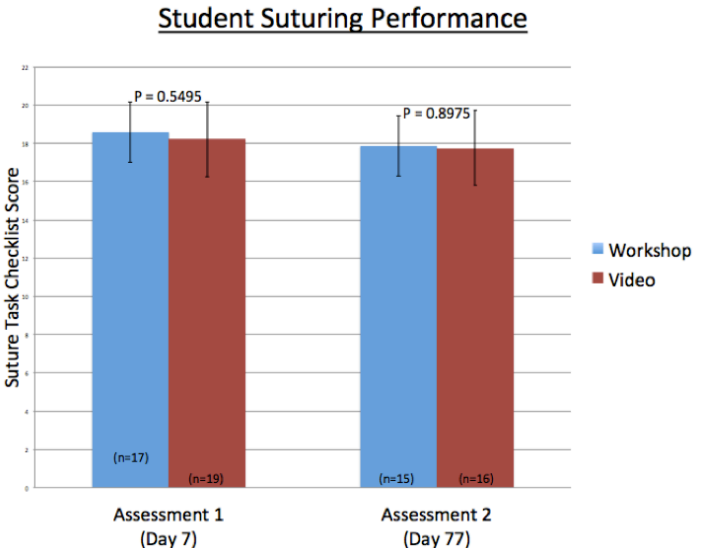


Figure 3. Demonstrates success of suturing performance based on 22-point check list as evaluated on days 7 and 77.

some limitations to this mode of learning include limited interaction with residents and physicians and lack of instructor feedback. Limitations to this study include not including baseline/pre-intervention evaluation of subjects suturing skillsets, small sample size, and the quality of the overhead camera used to capture suturing techniques and ties may not have been optimal for high resolution viewing at home. Future studies may look to evaluate VBL performance beyond a controlled practice environment and into real-life clinical situations. In addition, student preference between VBL and LWL should be assessed.

Grant Citation

The project described was supported by the National Center for Advancing Translational Sciences, through Grant UL1TR000117. The content is solely the responsibility of the authors and does not necessarily represent the University.

Address for Correspondence: Terren Trott, MD, University of Kentucky, 222 Arlington Ave. Lexington, KY 40508. Email: terren.trott@uky.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Khan MS, Bann SD, Darzi AW, et al. Assessing Surgical Skill Using Bench Station Models. *Plast Reconstr Surg*. 2007;120(3):793-800.
2. Conner HM and McGraw RC. Clinical skills training: developing objective assessment instruments. *Med Educ*. 1997;31(5):359-63.
3. Malon M. "Medical students' assessment of pediatric patients – teaching and evaluation using video cases." *BMC Med Educ*. 2014. 14:241.

Coordinating a Team Response to Behavioral Emergencies in the Emergency Department: A Simulation-Enhanced Interprofessional Curriculum

Ambrose H. Wong, MD*
Lisa Wing, RN†
Brenda Weiss‡
Maureen Gang, MD§

*Yale University School of Medicine, Department of Emergency Medicine, New Haven, Connecticut
†Bellevue Hospital Center, Adult Emergency Services, New York, New York
‡Bellevue Hospital Center, Hospital Police, New York, New York
§New York University School of Medicine, Department of Emergency Medicine, New York, New York

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted April 10, 2015; Revision received August 4, 2015; Accepted August 18, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.26220

Introduction: While treating potentially violent patients in the emergency department (ED), both patients and staff may be subject to unintentional injury. Emergency healthcare providers are at the greatest risk of experiencing physical and verbal assault from patients. Preliminary studies have shown that a team-based approach with targeted staff training has significant positive outcomes in mitigating violence in healthcare settings. Staff attitudes toward patient aggression have also been linked to workplace safety, but current literature suggests that providers experience fear and anxiety while caring for potentially violent patients. The objectives of the study were (1) to develop an interprofessional curriculum focusing on improving teamwork and staff attitudes toward patient violence using simulation-enhanced education for ED staff, and (2) to assess attitudes towards patient aggression both at pre- and post-curriculum implementation stages using a survey-based study design.

Methods: Formal roles and responsibilities for each member of the care team, including positioning during restraint placement, were predefined in conjunction with ED leadership. Emergency medicine residents, nurses and hospital police officers were assigned to interprofessional teams. The curriculum started with an introductory lecture discussing de-escalation techniques and restraint placement as well as core tenets of interprofessional collaboration. Next, we conducted two simulation scenarios using standardized participants (SPs) and structured debriefing. The study consisted of a survey-based design comparing pre- and post-intervention responses via a paired Student *t*-test to assess changes in staff attitudes. We used the validated Management of Aggression and Violence Attitude Scale (MAVAS) consisting of 30 Likert-scale questions grouped into four themed constructs.

Results: One hundred sixty-two ED staff members completed the course with >95% staff participation, generating a total of 106 paired surveys. Constructs for internal/biomedical factors, external/staff factors and situational/interactional perspectives on patient aggression significantly improved ($p<0.0001$, $p<0.002$, $p<0.0001$ respectively). Staff attitudes toward management of patient aggression did not significantly change ($p=0.542$). Multiple quality improvement initiatives were successfully implemented, including the creation of an interprofessional crisis management alert and response protocol. Staff members described appreciation for our simulation-based curriculum and welcomed the interaction with SPs during their training.

Conclusion: A structured simulation-enhanced interprofessional intervention was successful in improving multiple facets of ED staff attitudes toward behavioral emergency care. [West J Emerg Med. 2015;16(6):859–865.]

INTRODUCTION

Patients with behavioral emergencies often present to the emergency department (ED) in acute agitation.[1] The etiology of their agitation carries a wide differential diagnosis, including metabolic derangements, intracranial pathology, toxic and illicit drug ingestions and psychiatric emergencies.[2] Care of these patients comes with innate safety risks for both the staff members and the patients themselves. EDs have been identified as high-risk settings for workplace violence (WPV).[3] A national survey of emergency physicians found that at least one WPV act was reported in 78% of all responders, and 21% reported more than one episode.[4] Several studies have shown that emergency nurses are at the greatest risk of experiencing verbal and physical assault as compared to nurses in other healthcare settings and to physicians.[5,6] Precipitators of violence and aggression in the ED are most commonly attributed to alcohol/substance abuse, mental illness, and altered perceptions and confusion, all of which frequently exist in agitated patients.[7,8] Management of acutely agitated patients consists not only of physical restraints and administration of appropriate medications but also utilization of de-escalation and agitation reduction techniques.[2,9]

Recent surveys of healthcare workers have identified a need for early communication of clear roles and responsibilities of hospital security and ED staff to improve safety during WPV events.[10] Implementation of a structured team approach that promotes interprofessional collaboration to manage patients with behavioral emergencies has shown significant impact on mitigating aggression.[11,12] In addition, improving providers' attitudes and comprehension of factors contributing to patient violence has been directly linked to an improved workplace safety climate.[13] However, staff members have expressed ongoing fear and anxiety when caring for potentially aggressive patients, even to the point where some providers intentionally avoided engaging patients and visitors whom they deemed to have violent tendencies in order to alleviate their stress symptomatology.[14,15] Currently, educational strategies targeting WPV, including the widely adopted Nonviolent Crisis Intervention program from the Crisis Prevention Institute (CPI), have focused on an individual provider's interaction with violent patients.[16,17] Healthcare simulation provides a realistic but safe venue to address issues surrounding patient violence. More importantly, simulation–based education can both directly influence participant attitudes and encourage interprofessional teamwork due to its inherent ability to impact learners' cognitive frames and promote peer-to-peer dialog during structured debriefing.[18-20]

Our study used a multi-modality, team-based approach to create a novel simulation-enhanced patient safety curriculum targeting staff attitudes toward patient aggression and interprofessional collaboration during the management of patients with behavioral emergencies in the ED. We assessed the potential success of the program through direct analysis of staff attitudes towards management of aggression with a validated survey instrument. Our hope was that this intervention would allow for a coordinated team approach that would improve safety for both patients and staff members.

Assessment Instrument

We examined changes to staff attitudes as a result of this intervention via the Management of Aggression and Violence Attitude Scale (MAVAS), a published survey from a British nursing education group that has shown reliability and internal validity for assessment of staff attitudes toward patient aggression.[21] Although this survey's validation process was performed with a psychiatric patient cohort, transferability to the ED environment is feasible given that ED patients presenting with behavioral emergencies often carry psychiatric etiologies. Moreover, the survey authors included the psychiatric ED as one of the clinical environment for their investigation and thus allowed for applicability to the ED setting.[22] The survey was subdivided into four constructs contributing to patient aggression: internal and biomedical factors of the patient; external and staff factors; situational/interactional perspectives; and staff perspectives toward management of patient aggression. This was distributed immediately pre- and post-session to assess the course's direct impact on our staff. See Appendix 1 for a copy of the survey questions and elements. We used the paired sample Student's *t*-test for our survey data analysis using IBM SPSS 21.0 software, and our study was approved by NYU School of Medicine's Institutional Review Board as expedited review under the title, "Simulation-based Team Training for Care of Acutely Agitated Patients in the Emergency Department (i14-00846)" in May 2014.

EDUCATIONAL OBJECTIVES

1. Describe and demonstrate effective interprofessional teamwork and communication skills to treat the patient with a behavioral emergency in the ED.
2. Identify roles and responsibilities of members of an interprofessional team that care for acutely agitated patients.
3. Display effective violence mitigation and de-escalation techniques.
4. Appropriately apply physical restraints and medical interventions during treatment of the agitated patient in the ED.
5. Demonstrate improvement in attitudes toward patients with behavioral emergencies through a better understanding of factors contributing to patient aggression.

CURRICULAR DESIGN

As the care of the agitated patient requires balancing a complex range of clinical, communication and teamwork skills, we felt that applying David Kolb's experiential learning theory as our educational framework would best suit our needs.[23] We developed a simulation-enhanced

interprofessional curriculum as the application of experiential learning where our physicians, nurses, patient care technicians and hospital police officers trained together to replicate the ED clinical environment. Content experts from the hospital's Crisis Management Team (CMT) with extensive training in teaching de-escalation techniques and evidence-based management of aggressive persons joined us in our educational endeavor. They worked with physician, nursing and hospital police educational leadership to ensure that our curricular content was in line with current best practices from the literature.[2,3,17] We used standardized participants (SPs) to maximize fidelity during case-based simulations that were designed to incorporate de-escalation and personal defense techniques, team-based interprofessional approaches, application of physical restraints and adjunctive medication route and dosing options.

Didactics

Our educational team derived a 30-minute introductory interactive lecture from core elements of validated aggression management courses. Key components of the didactics including crisis management principles, de-escalation techniques, and proper application of restraints were summarized in a pre-session handout that was distributed to our learners prior to beginning of the session. Moreover, ED leadership constructed formal roles and responsibilities prior to the didactic session. These roles were described in detail in the pre-session handout and re-enforced with staff during the didactic session (see Figure 1 for detailed description). At the end of the didactic component, we solicited participating staff for quality improvement initiatives that could be implemented in the clinical setting to further advance staff and patient safety during treatment of the patient with a behavioral emergency.

Immersive Simulated Encounters

We recruited healthcare professionals to act as SPs and trained them in conjunction with our content experts to simulate two agitated patient scenarios typically encountered by ED staff members. The course participants were expected to use the de-escalation techniques in an interprofessional manner discussed during the didactic session to calm the simulated agitated patients. The simulations were designed so that de-escalation techniques would only be partially successful, and the team would then need to apply physical restraints and medical therapy to complete the scenarios. A code phrase, "mickey mouse," was designated as a "time out" should participants or SPs feel that they were in physical danger or out of their comfort zone during the scenarios, while the educator team closely monitored each simulation encounter from the control room. At the completion of each immersive simulation, the interprofessional group of participants immediately proceeded to a structured debriefing session led by health professions educators specifically trained in educational theory and debriefing concepts. We ensured that the main discussion points focused on participant attitudes towards factors contributing to patient aggression as well as interprofessional collaboration and communication skills demonstrated during the encounters.

Team leader (attending or senior resident physician)
- Assigns roles clearly at EMS notification of patient or on patient arrival
- Removes non-essential personnel, controls traffic
- Stays at foot of bed and does not become involved with procedures/assessment unless confirmation of abnormal finding needed
- Communicates with nursing regarding restraint placement and medications, enforces closed loop communication for all orders
- Monitors overall safety of patient and clinical providers during encounter

"Patient point person" (any clinical provider)
- Primary person that has the best rapport with the patient
- Communicates directly with patient during the initial encounter and restraint/medication process (if clinically necessary)
- Does not apply any restraints
- May change to another provider as alliance with patient changes
- Monitors patient's airway and breathing status

Clinical staff member – restraint placement (one per limb of patient)
- Undresses and covers patient
- Coordinates with police officers to safely place restraints on patient
- Calls out "[limb] secure!" for the limb responsible
- Monitors the safety of the patient

Clinical staff member - medical (if available)
- Retrieves and draws up medications as clinically necessary
- Places patient on monitor
- Records vital signs and condition of patient
- Places IV and obtains blood work once it is safe and feasible to do so (with orders from team leader)
- Performs physical exam

Hospital police officer (at least one per limb of patient)
- Stabilizes patient's extremities while restraints are being applied
- Records encounter into police records as appropriate
- Assists in crowd control and patient privacy

Figure 1. Agitated patient care team: roles and responsibilities. This is set with the model of 2-3 nurses, 2-3 physicians, and 2-3 police officers, and 1-2 ancillary staff members.
EMS, emergency medical services; *IV*, intravenous

Case 1: Intoxicated Patient with Head Trauma

The first scenario involved a patient who was brought to the ED by paramedics for evaluation of altered mental status and minor head trauma. The patient appeared to be intoxicated with alcohol and became angry and threatening during the triage process. The participants were required to use the de-escalation techniques demonstrated during the didactic session, recognize

that the patient had head trauma as evidenced by a scalp laceration, order a head computed tomography and diagnose a subdural hematoma that was due to an acute fall from a standing position. Discussions focused on workflow and restraint placement as a large consumption of manpower and resources, especially on weekend overnight shifts when many intoxicated patients presented simultaneously in the ED.

Case 2: Psychiatrically ill patient with sympathomimetic toxidrome

The second scenario featured a physically and verbally aggressive patient with underlying psychiatric illness who initially responded to de-escalation by staff. However, he quickly became more aggressive and dangerous despite participants' attempts. He was later found to have ingested phencyclidine, requiring medical therapy and safe restraint placement. Facilitators often needed to intervene during this second simulation to halt the scenario and correct participant errors, using a strategy similar to "rapid cycle deliberate practice" training.[24] We noticed potential real physical danger to the participants or the SP due to the physical nature of the case. Staff often raised concerns regarding the durability of the restraints and specific mechanical details of the restraint placement process to prevent injuries, which were clarified by our educators and CMT experts.

Implementation Strategies

Engaging and securing administrative support was key to the successful implementation of our intervention. To minimize disruption of clinical care, sessions were incorporated within already established training time periods for nurses and resident physicians. For nursing, we incorporated this course into their annual competency training. Simulation didactics for the residents were scheduled on a weekly basis in the simulation center and 10 of those sessions were used for this course.

IMPACT

Survey Results & Staff Response

In total, we conducted 10 three-hour sessions from July to September 2014. One hundred sixty-two ED staff members completed the course with >95% staff participation, generating a total of 106 paired pre-post surveys. See the Table for a detailed list of survey respondent demographics. Constructs for internal factors, external factors and situational/interactional perspectives on patient aggression significantly improved post-intervention ($p<0.0001$, $p<0.002$, $p<0.0001$ respectively, Figure 2). Staff attitudes toward management of patient aggression did not significantly change ($p=0.542$). Secondarily, staff participants gradually generated a list of quality improvement initiatives as the weeks went by, many of which were successfully implemented including the creation of an ED-based interprofessional crisis management alert and response protocol.

The results of the MAVAS survey reflected our staff participants' immediate changes in attitudes toward patient aggression factors as a result of our course except in the construct of clinical management of aggression. As our curriculum objectives focused heavily on prevention and recognizing factors contributing to aggression rather than the specific medical management of aggression, the survey accurately reflected our intended interventions. In fact, we wished to deliberately not discuss details of clinical decision-making while caring for our targeted population for the purposes of this course. Our agitated ED patients present with a breadth of medical and psychiatric etiologies, and management depends heavily upon the unique circumstances and ultimate diagnoses of a particular patient encounter.

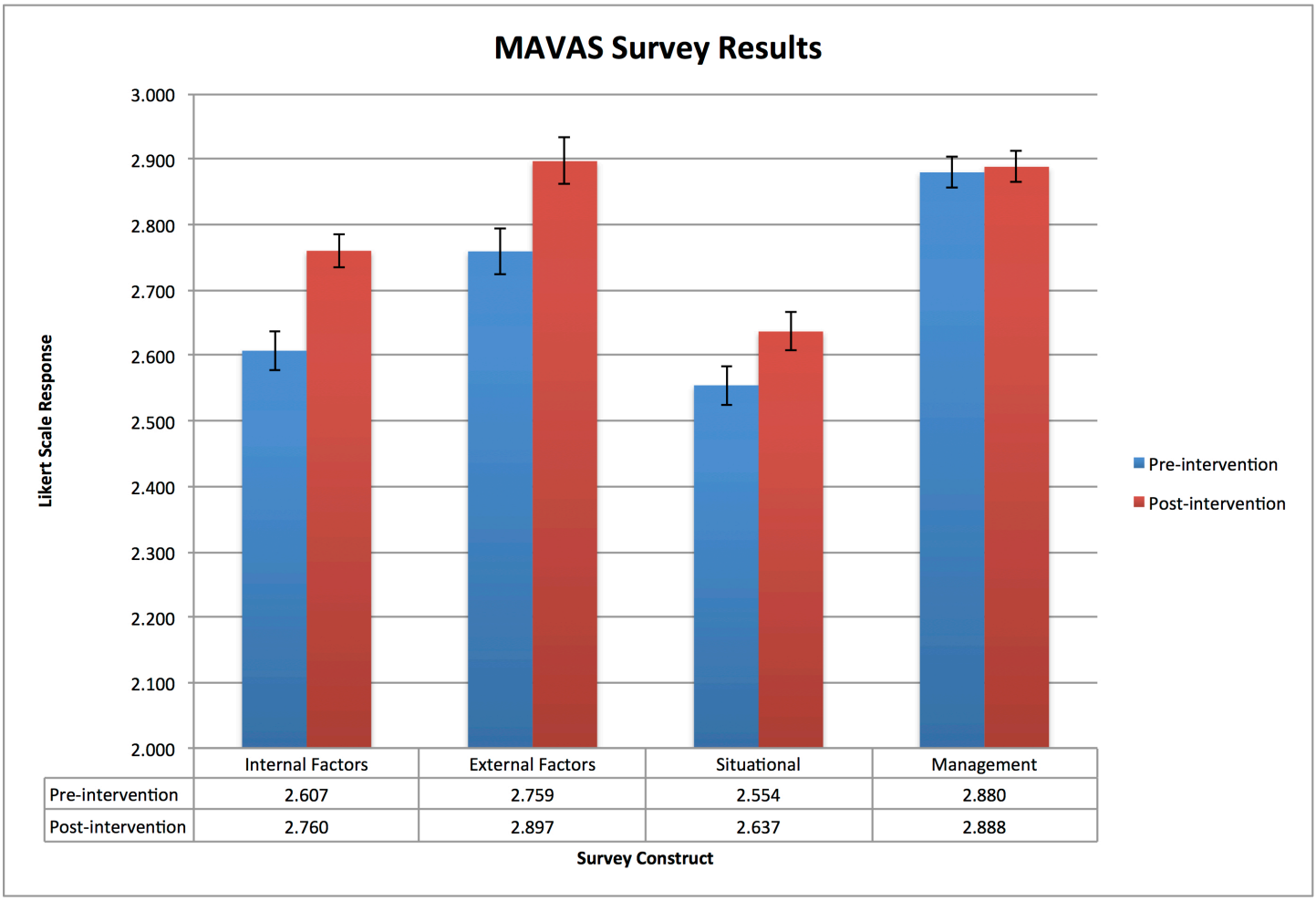
Staff participants overwhelmingly endorsed and welcomed the SPs in the hands-on components of the course and frequently commented on how having SPs in the simulations significantly increased fidelity and helped recreate a realistic scenario for them. Many in fact forgot that they were participating in a simulation altogether and experienced the same fear, anxiety and frustrations that they felt while caring for an agitated patient in a prior clinical shift. Although none of the SPs or learners used the "time out" even once or suffered any injuries during the course, we paid close attention to the play of the scenarios in the control room to observe for latent safety lapses. As mentioned above, educators entered the simulation room on multiple occasions to pause the scenario and intervene in the second case with the acutely aggressive patient.

Pitfalls and Limitations

This course was time and resource intensive for the instructors. Even with an average of 15-20 learners per session, we required 10 sessions to completely train our department. Each session required at least nine to ten instructors and assistants to run the simulations and lead the interprofessional debriefing sessions. We found that having a core team of

Table. Survey respondent demographics.

Characteristic	N
Staff clinical role	
Ancillary staff	6
Nurse	43
Physician	36
Hospital police	21
Gender	
Male	44
Female	62
Age group	
21 to 25	2
26 to 30	36
31 to 35	14
36 to 40	13
41 to 45	12
46 to 50	9
51 to 55	11
56 or older	9



	Internal Factors	External Factors	Situational	Management
Pre-intervention	2.607	2.759	2.554	2.880
Post-intervention	2.760	2.897	2.637	2.888

Figure 2. Management of Aggression and Violence Attitude Scale (MAVAS) survey results showing emergency department personnel's changes in attitude to patient aggression after participation in course on managing agitated patients.

nursing, physician and police educators, at least one of whom was trained in immersive simulation and debriefing structure and techniques, was critical to ensuring continuity and consistency between sessions. Our program benefited from the availability of a robust simulation center at our institution. We believe that other educators and administrators interested in implementing this curriculum can still successfully conduct the simulations and debriefing sessions in small meeting rooms or other office spaces within a hospital or learning environment. With regards to program evaluation, the cohort of educators and researchers for our pilot study were also in leadership positions within the department, which may have confounded our participants' responses to the MAVAS survey.

Future Directions

Additional work includes longitudinal data collection of staff attitudes over longer time periods, comparison of different methods of training and curriculum design, as well as a higher level of evaluation in the definitions of translational educational research to include patient outcomes or direct indices of care safety and quality.[18,25] Finally, validation studies of our interprofessional curriculum across different clinical sites may expand the applicability of the training methodology used in our study to a wider spectrum of institutions and departments.

In order to promote sustainability, stricter implementation of the defined roles and quality improvement initiatives need to occur on a consistent basis with buy-in from administration and staff members across all professions. A qualitative analysis of ongoing barriers and staff concerns to sustaining these efforts and caring for our agitated patient population may bring more key issues to light. Finally, the curriculum can be re-enforced with repeat sessions at scheduled intervals with shorter didactics or targeted to new staff hires and incoming physician trainees.

CONCLUSION

An interprofessional simulation-based team-training curriculum successfully improved staff attitudes toward the factors impacting the care of patients with behavioral emergencies in the ED. We hope the next steps in interprofessional education research will lead us toward sustainable and outcomes-based measures to improve patient and staff safety utilizing team effectiveness in caring for the

potentially aggressive patient.

ACKNOWLEDGEMENTS

The research was supported by the New York University School of Medicine's Program for Medical Education Innovations and Research (PrMEIR) 2014-2015 grant cycle. The authors would like to thank our contributors Dr. Tsion Firew, Dr. Neel Naik, Dr. Kelly Medwid, Antoinette Cirillo, Joyce Buffolino, Rob Rosenwald and Captain George Ortega for their work on this project. Additionally, we would like to thank the support staff at the New York Simulation Center for the Health Sciences (NYSIM), leadership of Bellevue Hospital Adult Emergency Services, and the ED physicians, nurses, clerks and patient care technicians for their dedication to providing safer care and for their assistance in the completion of this project. Finally, we would like to thank Dr. Joy Duxbury for allowing us to use the MAVAS survey as part of our study and for sharing her materials for the purposes of this publication.

Address for Correspondence: Ambrose H. Wong, MD, Yale University School of Medicine, 464 Congress Avenue Suite 260, New Haven, CT 06519. Email: ambrose.wong@yale.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Bazarian JJ, Stern RA, Wax P. Accuracy of ED triage of psychiatric patients. *Am J Emerg Med.* 2004;22:249-53.
2. Downey LV, Zun LS, Gonzales SJ. Frequency of alternative to restraints and seclusion and uses of agitation reduction techniques in the emergency department. *Gen Hosp Psychiatry.* 2007;29:470-4.
3. Kowalenko T, Cunningham R, Sachs CJ, et al. Workplace violence in emergency medicine: current knowledge and future directions. *J Emerg Med.* 2012;43:523-31.
4. Behnam M, Tillotson RD, Davis SM, et al. Violence in the emergency department: a national survey of emergency medicine residents and attending physicians. *J Emerg Med.* 2011;40:565-79.
5. Hodge AN and Marshall AP. Violence and aggression in the emergency department: a critical care perspective. *Aust Crit Care.* 2007;20:61-7.
6. Wolf LA, Delao AM, Perhats C. Nothing Changes, Nobody Cares: Understanding the Experience of Emergency Nurses Physically or Verbally Assaulted While Providing Care. *J Emerg Nurs.* 2014;40:305-10.
7. van der Zwan R, Davies L, Andrews D, et al. Aggression and violence in the ED: issues associated with the implementation of restraint and seclusion. *Health Promot J Austr.* 2011;22:124-7.
8. Arnetz JE, Hamblin L, Essenmacher L, et al. Understanding patient-to-worker violence in hospitals: a qualitative analysis of documented incident reports. *J Adv Nurs.* 2015;71:338-48.
9. Richmond JS, Berlin JS, Fishkind AB, et al. Verbal De-escalation of the Agitated Patient: Consensus Statement of the American Association for Emergency Psychiatry Project BETA De-escalation Workgroup. *West J Emerg Med.* 2012;13:17-25.
10. Gillespie GL, Gates DM, Miller M, et al. Emergency department workers' perceptions of security officers' effectiveness during violent events. *Work.* 2012;42:21-7.
11. Rintoul Y, Wynaden D, McGowan S. Managing aggression in the emergency department: promoting an interdisciplinary approach. *Int Emerg Nurs.* 2009;17:122-7.
12. Downes MA, Healy P, Page CB, et al. Structured team approach to the agitated patient in the emergency department. *Emerg Med Australas.* 2009;21:196-202.
13. Gerdtz MF, Daniel C, Dearie V, et al. The outcome of a rapid training program on nurses' attitudes regarding the prevention of aggression in emergency departments: a multi-site evaluation. *Int J Nurs Stud.* 2013;50:1434-45.
14. Gillespie GL, Gates DM, Berry P. Stressful incidents of physical violence against emergency nurses. *Online J Issues Nurs.* 2013;18:2.
15. Luck L, Jackson D, Usher K. Innocent or culpable? Meanings that emergency department nurses ascribe to individual acts of violence. *J Clin Nurs.* 2008;17:1071-8.
16. Calabro K, Mackey TA, Williams S. Evaluation of training designed to prevent and manage patient violence. *Issues Ment Health Nurs.* 2002;23:3-15.
17. Gillam SW. Nonviolent crisis intervention training and the incidence of violent events in a large hospital emergency department: an observational quality improvement study. *Adv Emerg Nurs J.* 2014;36:177-88.
18. Palaganas JC, Epps C, Raemer DB. A history of simulation-enhanced interprofessional education. *J Interprof Care.* 2014;28:110-5.
19. Kenaszchuk C, MacMillan K, van Soeren M, et al. Interprofessional simulated learning: short-term associations between simulation and interprofessional collaboration. *BMC Med.* 2011;9:29.
20. Aston SJ, Rheault W, Arenson C, et al. Interprofessional education: a review and analysis of programs from three academic health centers. *Acad Med.* 2012;87:949-55.
21. Duxbury J. Testing a new tool: the Management of Aggression and

Violence Attitude Scale (MAVAS). *Nurse Res.* 2003;10:39-52.

22. Duxbury J, Hahn S, Needham I, et al. The Management of Aggression and Violence Attitude Scale (MAVAS): a cross-national comparative study. *J Adv Nurs.* 2008;62:596-606.
23. Kolb D. *Experiential learning : experience as the source of learning and development.* 2nd edition. ed. Indianapolis, IN: Pearson; 2014.
24. Hunt EA, Duval-Arnould JM, Nelson-McMillan KL, et al. Pediatric resident resuscitation skills improve after "rapid cycle deliberate practice" training. *Resuscitation.* 2014;85:945-51.
25. McGaghie WC, Issenberg SB, Cohen ER, et al. Translational educational research: a necessity for effective health-care improvement. *Chest.* 2012;142:1097-103.

Development of an Objective Structured Clinical Examination for Assessment of Clinical Skills in an Emergency Medicine Clerkship

Sharon Bord, MD
Rodica Retezar, MD, MPH
Pamela McCann
Julianna Jung, MD

Johns Hopkins University School of Medicine, Department of Emergency Medicine, Baltimore, Maryland

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted May 1, 2015; Revision received September 28, 2015; Accepted September 29, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27307
[West J Emerg Med. 2015;16(6):866–870.]

INTRODUCTION

Assessment of medical students in their emergency medicine (EM) clerkship is often based on clinical shift evaluations and written examinations. Clinical evaluations offer some insight into students' ability to apply knowledge to clinical problems, but are notoriously unreliable, with score variance that may be driven as much by error as by actual student performance.[1-6] Clinical evaluations are also limited by the unpredictability of pathology in emergency department (ED) patients, and by patient safety considerations that prevent students from independently managing patients, especially those with high-acuity conditions. Additionally, there is evidence that the basic skills of history and physical exam are rarely observed by faculty members, and the feedback they receive on these domains is limited.[7-9] These factors hinder EM educators in their effort to objectively assess students' progress relative to clerkship objectives, particularly those that pertain to emergent care.

The objective structured clinical exam (OSCE) is one potential solution to these problems. Described in 1975 by Harden et al, an OSCE is designed to assess the clinical competence of medical trainees through direct observation of skill performance in a variety of stations.[10] OSCEs have been widely adopted in medical education and in other health professions.[11-13] These exams are viewed as a valuable form of clinical assessment due to their demonstrated reliability and inherent flexibility for assessing a wide variety of knowledge application and skills.[11]

In EM, OSCEs have been used mainly in postgraduate medical education to assess resident communication skills and clinical performance.[14,15] One OSCE for interns was shown to accurately predict future clinical performance scores.[12] A recently published evaluation of an EM OSCE for medical students demonstrated validity evidence supporting this method of assessment.[16] These studies suggest that OSCEs can be used effectively in EM education, and yield valid assessment data.

OBJECTIVE

Our primary goal was to develop an OSCE that would assess whether students have not only acquired essential knowledge during their EM clerkship, but are also able to synthesize this knowledge into a management plan and perform key critical actions in emergency situations. EM is a required core clerkship for all students at our institution, and occupies an essential place our medical student curriculum. While many students will not pursue careers in EM, all will confront emergencies throughout their careers, and emergency management skills are vital for all physicians. Rigorous assessment will allow students to appropriately focus their future learning to improve their skills in this arena, it will facilitate curricular improvements by giving educators insight into common errors and misconceptions, and it will permit documentation of student competency.

CURRICULAR DESIGN

We designed our OSCE for use in the required core clerkship in EM, which is taken by students ranging from the final quarter of the MS-2 year to the third quarter of the MS-4 year. The exam is fundamentally criterion-referenced, in that all material covered in the exam is explicitly taught during the clerkship, and it is our expectation that all students will "pass." However, as is the case with many assessments in medical education, there is an element of norm-referencing as well. While it was our hope that every student will demonstrate at least minimal competency, we expected to

see a wide range of exam performance based on variation in student knowledge and abilities, and we sought to capture that range when developing the exam.

We developed the OSCE based on the EM Milestones, with the goal of assessing the majority of the clinically-oriented competencies described in that framework. The milestones that are assessed with exam include emergency stabilization, performance of history and physical exam, diagnostic skills, diagnosis, pharmacotherapy, airway management, observation and reassessment, goal directed focused ultrasound, professional values and patient centered care (Table 1). The exam consists of three stations: two manikin-based simulations and one standardized patient encounter. These scenarios were developed and vetted by a core group of undergraduate medical educators in the department of EM at our institution. Specific cases and complaints were chosen in order to highlight challenges and topics that are unique to the field of EM and represented in the Milestones. All cases were designed to assess the student's ability to both assess and manage critical illness independently.

The cases depict the following diagnoses: massive pulmonary embolus, intra-abdominal bleeding after blunt trauma, and poly-substance overdose with airway compromise. Each case has two phases: a stable "assessment" phase, and an unstable "treatment" phase. During the student's evaluation, all patients become unstable, necessitating rapid resuscitation. Cases are described below, and key checklist items are summarized and mapped to Milestones in Table 1:

Case 1: Massive pulmonary embolus. In this manikin scenario, a patient presents three days following total knee replacement with acute chest pain and dyspnea. The patient develops pulseless electrical activity that later degenerates to ventricular fibrillation. Students are expected to perform needed assessment, order appropriate diagnostics for pulmonary embolus, and provide advanced cardiovascular life support interventions for cardiac arrest.

Case 2: Blunt abdominal trauma. In this standardized patient encounter, a young female patient is assaulted and presents with head, neck, and abdominal pain. She complains of worsening global weakness and dizziness throughout the encounter as she develops hemorrhagic shock from intra-abdominal bleeding. Students are expected to perform an appropriate primary and secondary survey, initiate needed volume resuscitation, request focused abdominal sonography, and consult surgery.

Case 3: Poly-substance overdose. In this manikin scenario, the patient presents unconscious and is not able to give a history. Examination reveals empty pill and liquor bottles in his pockets. The patient becomes progressively less responsive during the encounter and develops hypoxemia due to airway obstruction and respiratory depression. Students are expected to perform needed assessment, order appropriate diagnostics for altered mental status, and provide airway and respiratory management as hypoxia worsens.

Prior to the OSCE, the students receive a standardized orientation to exam procedures and logistics. During the course of the EM clerkship, students have approximately 20 hours of instructional time, all of which is simulation-based. They are therefore very familiar with simulation and comfortable in the simulation environment prior to the exam.

Following a brief introduction to the patient's presenting complaint, the student enters and begins the case. Each station includes an in-room confederate playing the role of the patient's nurse. The confederate roles are extensively scripted to ensure standardization. Confederates are permitted to assist students with locating equipment, obtaining clinical data, administering medications, and performing limited clinical interventions (e.g., chest compressions). Students must otherwise be self-sufficient, as confederates are not permitted to offer suggestions about diagnosis or treatment, and are not permitted to perform clinical interventions outside of their limited scope. Confederates do provide standardized prompts to ensure that the case proceeds in an expeditious fashion, though actions that are prompted by confederates are not given credit on the scoring checklist. The confederates are usually EM educators, though may occasionally be portrayed by simulation center staff members.

All stations are 10 minutes in length with an additional two minutes provided for student feedback. A simulation center staff member controls the timing of the examination. Students receive brief feedback on their performance from the observing faculty member and standardized patient following each case. To ensure psychological safety of the learners, case conclusions are standardized, with all patients stabilized by the end of the scenario. Confederate prompting is designed to ensure that students complete all "life and death" actions, enabling successful resuscitation of the patient before the scenario ends.

Following each station, students are graded using a structured checklist completed in real time by an observing EM faculty member. The standardized patient also completes a checklist to evaluate the students on history-taking and interpersonal skills. Each checklist ranges in length from 29-37 items. The OSCE score is determined by calculating percent of checklist items completed correctly for each station. Station percentages are then averaged to determine the final score, in order to ensure that all stations are weighted equally. For exam security reasons, we are not able to append the final checklists, though we would be happy to privately share our test materials with other educators.

IMPACT/EFFECTIVENESS

During the first eight months of its administration, the OSCE was used as a pilot test and the students' performance did not count towards their final grade in the clerkship.

Table 1. Mapping checklist items to emergency medicine milestones.

Milestone	Description	Case(s)	Checklist items
1	Emergency stabilization	All	- Recognition of abnormal vital signs - Primary assessment on critically injured patient
2	Performance of focused history and physical exam	All	- Obtains focused history - Obtains focused physical examination
3	Diagnostic studies	All	- Requests CXR and ECG - Considers Chest CT - Requests appropriate laboratory testing (including acetaminophen level, alcohol level, d-dimer)
4	Diagnosis	All	- Considers pulmonary embolus in diagnosis - Considers intra-abdominal bleeding in diagnosis - Considers acetaminophen, opioid, alcohol overdose
5	Pharmacotherapy	All	- Requests and administers rapid sequence intubation medications - Administers epinephrine in cardiac arrest - Requests N-acetyl cysteine for treatment of acetaminophen overdose
6	Observation and reassessment	All	- Reassesses vital signs after return of spontaneous circulation following cardiac arrest - Reassesses vital signs after administration of intravenous fluids
10	Airway management	3	- Effectively bag valve masks patient and troubleshoots BVM technique - Performs rapid sequence intubation in patient with airway compromise
12	Goal-directed focused ultrasound	2	- Orders FAST exam in patient with abnormal vital signs following trauma
20	Professional values	2	- Standardized patient assesses student's interpersonal skills - Demonstrates behavior that conveys caring, honesty and genuine interest and tolerance
22	Patient centered care	2	- Standardized patient assesses student's interpersonal skills - Establishes rapport and demonstrates empathy towards patient - Effectively listens to patient

CXR, chest x-ray; *ECG*, electrocardiogram; *CT*, computed tomography; *FAST*, focused assessment with sonography for trauma

Following the pilot period, we analyzed the OSCE data and revised the exam and curriculum accordingly. Currently, performance on the OSCE represents 20% of the student's final clerkship grade. Other grade components include daily clinical evaluations (55%), a direct observation session (5%), and an internally developed written exam (20%). While we have labored to make our OSCE as psychometrically sound as possible, we recognize the inherent reliability limitations of an exam with a small number of stations. We therefore elected to make it a relatively small part of the students' final grades.

The OSCE pilot period included a convenience sample of 80 students, all of whose performance data was analyzed. The average score on the exam was 70.5%, with a standard deviation of 7.2%. Scores ranged from 39.3 to 84.1%, and grade distribution data are detailed in the Figure. Of note, the OSCE offers the widest grade distribution of any assessment method used in our clerkship, allowing us to effectively discern students who excel from those who struggle.

Item analysis was completed for all 96 checklist items, including difficulty and discrimination values. Average item difficulty was 70.0%, which is in the "medium" range and is considered appropriate. Average point biserial correlation (r_{pb}) was 0.24, which is in the "fair" range, and is considered acceptable though not ideal.

Item difficulty results are presented in Table 2. While there are no universal definitions of item difficulty in educational research, the cutoffs we selected are common. High-difficulty items may suggest a problem with the case, in that it does not provide sufficient clinical clues to prompt students to complete desired actions, or with the items themselves, in that they are not clinically relevant to the case as presented. These items may also reflect a problem with the curriculum, in that the desired action is not being adequately taught. High-difficulty items should not form the foundation on an exam, but some of these items are necessary and desirable for differentiating between low-

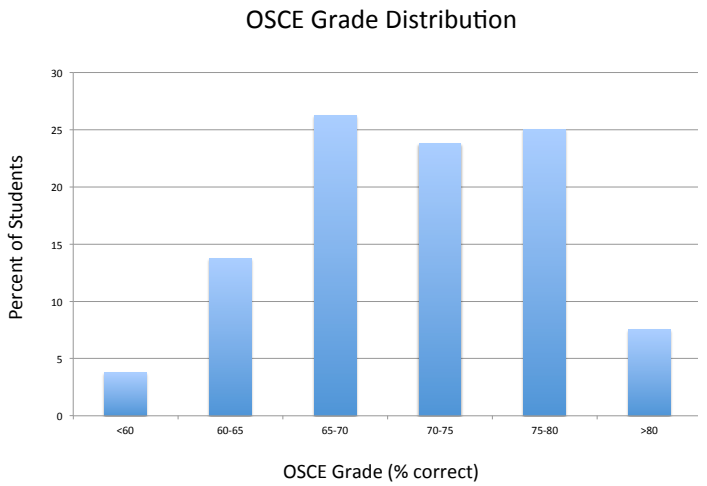


Figure. Objective structured clinical exam (OSCE) grade distribution. Each bar represents the percent of students achieving a final score within the grade range listed on the X-axis.

Table 2. Item difficulty of OSCE components.

Difficulty level	Number of checklist items
Low (>80% correct)	36 (37%)
Medium (50-80% correct)	41 (43%)
High (<50% correct)	19 (20%)

Table 3. Item discrimination of OSCE components.

Discrimination level	Number of checklist items
Good (r_{pb}>0.3)	10 (11%)
Fair (r_{pb}=0.1-0.3)	47 (51%)
Poor (r_{pb}<0.1)	35 (38%)

and high-performing students, as it would be expected that only top students would get these items correct. Likewise, low-difficulty items should also not be overrepresented on an exam, but some of these items are appropriate for the documentation of critical, foundational competencies that every student is expected to know.

Item discrimination results are presented in Table 3. Poor discrimination means that overall low-performing learners get an item correct, while high performers get the item incorrect. Easy items will always discriminate poorly, as they are completed correctly by low-performers and high-performers alike. As noted above, this is not always problematic, particularly for items that reflect universal basic competencies, like initiating chest compressions for cardiac arrest. However, poor discrimination may also suggest a problem with clinical aspects of the case or item.

When completing an item analysis, difficulty and discrimination must be considered simultaneously, and item revision decisions are never made based solely on one or the other. For example, a high-difficulty item might be retained if it discriminates effectively between students. It is also essential to consider each item in the context of its purpose in the exam. For example, poor discrimination is acceptable for a foundational item that is expected to be "easy." However, it is problematic for an item that is intended to be difficult, as this means that strong and weak students are equally likely to miss the item, suggesting that it is either not taught in the curriculum or not adequately cued by the case.

Based on our item analysis, we removed 11 items. Of these, half were high-difficulty items that were poor discriminators, in most cases because they were too challenging for even strong students to complete in the very limited time available. The other half were low-difficulty items with poor discrimination that were not felt to be sufficiently foundational to remain on the list. We revised the case and/or item to address concerns about 13 items. Most revisions were made for high-difficulty items in which case was adjusted to make the need for the action in question more obvious. There were also six items that led to curricular adjustments in order to emphasize key points and teach key concepts more effectively.

LIMITATIONS

There are, of course, important limitations to address. First and foremost, this project was conducted at a single site, and the results may not generalize to other institutions. Second, we present only preliminary pilot data, though we are currently in process of implementing our revised exam with another group of students and will be able to determine whether our revisions improve the psychometric performance of the exam. Last and most important was our inability to fully validate the exam. Doing this would require development of a true "gold standard" against which to compare student performance on this assessment. No such standard currently exists, particularly for clinical performance. Given the very significant limitations of "real" clinical assessment, attainment of this standard may prove elusive.

That said, we believe that implementation of our OSCE has provided us with valuable information regarding the performance of our students at the completion of their basic clerkship. It gave us a unique window on their ability to independently evaluate and manage acutely ill and injured patients, as it required them to apply knowledge and skills gained throughout the course of the rotation to "real" clinical problems. We learned what things the students reliably do well, and where they struggle. We also identified common errors and misconceptions, enabling us to strengthen our teaching in these areas. This examination can easily be adapted at other institutions provided they have access to simulation technology. Our cases test core EM content reflected in clinically-oriented Milestones, and allow assessment of the student's ability to manage these issues

independently in a way that would be impossible in the clinical area.

CONCLUSION

Overall, we found that the OSCE effectively discriminates between high- and low-performing students in a way that other assessment tools do not. The score range on the OSCE is wider than that of our written exam, and our clinical evaluations (like those of many institutions) suffer from a severe restriction of range that limits their utility in differentiating between students – a problem we do not see with the OSCE. The OSCE also offers insight into aspects of student performance that are not captured through other means of evaluation. Overall, we believe that OSCEs offer a useful tool for assessment of EM knowledge and skills, and they can provide a foundation for documentation of the essential competencies reflected in the Milestones and the newer Entrustable Professional Activities.[17]

Address for Correspondence: Sharon Bord, MD, John Hopkins University School of Medicine, 1830 E. Monument St. 6th Floor, Baltimore, MD 21287. Email: sbord1@jhmi.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. McGill DA, van der Vleuten CP, Clarke MJ. Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Adv Health Sci Educ Theory Pract.* 2011;16(3):405-25.
2. McGill DA, van der Vleuten CP, Clarke MJ. A critical evaluation of the validity and the reliability of global competency constructs for supervisor assessment of junior medical trainees. *Adv Health Sci Educ Theory Pract.* 2013;18(4):701-25.
3. Daelmans HE, van der Hem-Stokroos HH, Hoogenboom RJ, et al. Global clinical performance rating, reliability and validity in an undergraduate clerkship. *Neth J Med*. 2005;63(7):279-84.
4. Carline JD, Paauw DS, Thiede KW, et al. Factors affecting the reliability of ratings of students' clinical skills in a medicine clerkship. *J Gen Intern Med*. 1992;7(5):506-10.
5. Kreiter CD and Ferguson KJ. Examining the generalizability of ratings across clerkships using a clinical evaluation form. *Eval Health Prof*. 2001;24(1):36-46.
6. Plymale MA, French J, Donnelly MB, et al. Variation in faculty evaluations of clerkship students attributable to surgical service. *J Surg Educ.* 2010;67(3):179-83.
7. Kuo AK, Irby DI, Loseser H. Does direct observation improve medical students' clerkship experiences? *Med Educ*. 2005;39(5):518.
8. 2012 AAMC Graduation Questionnaire Program Evaluation Survey: All Schools Summary Report Final. Available at: https://www.aamc.org/download/300448/data/2012gqallschoolssummaryreport.pdf. Accessed Sept 4, 2015
9. Burdick WP and Schoffstall J. Observation of emergency medicine residents at the bedside: how often does it happen? *Acad Emerg Med*. 1995;2(10):909-13.
10. Harden RM, Stevenson M, Downie W, et al. Assessment of clinical competence using objective structured examination. *BMJ*. 1975;1;447-51.
11. Chakravarthy B, Ter Haar E, Subraya Bhat S, et al. Simulation in medical student education: review for Emergency Medicine. *West J Emerg Med*. 2011;12(4):461-6.
12. Patricio MF, Juliao M, Faeleira, et al. Is OSCE a feasible tool to assess competencies in undergraduate medical education? *Med Teach.* 2013;35:503-14.
13. Rushforth HE. Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse Educ Today*. 2007;27(5):481-90.
14. Okuda Y, Bond W, Bonfante G, et al. National growth in simulation training within emergency medicine residency programs, 2003-2008. *Acad Emerg Med.* 2008;15:1113-6.
15. Wallenstein J, Heron S, Santen S, et al. A core-competency-based objective structured clinical examination (OSCE) can predict future resident performance. *Acad Emerg Med.* 2010;17:s67-71.
16. Wallenstein J and Ander D. Objective structured clinical examinations provide valid skills assessment in emergency medicine education. *West J Emerg Med*. 2015;16:121-6.
17. Chen HC, Van den Broek WE, Ten Cate O. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med*. 2015;90(4):431-6.

Direct Observation Assessment of Milestones: Problems with Reliability

Meghan Schott, MD*
Raashee Kedia, MD†
Susan B. Promes, MD*
Thomas Swoboda, MD‡
Kevin O'Rourke, MD§
Walter Green, MD¶
Rachel Liu, MD‖
Brent Stansfield, PhD#
Sally A. Santen, MD, PhD#

*University of California, San Francisco, Department of Emergency Medicine, San Francisco, California
†Icahn School of Medicine at Mount Sinai, Department of Emergency Medicine, New York, New York
‡Louisiana State University Health Sciences Center – Shreveport, Department of Emergency Medicine, Shreveport, Louisiana
§New York Methodist Hospital, Department of Emergency Medicine, Brooklyn, New York
¶Parkland Memorial Hospital, Department of Emergency Medicine, Dallas, Texas
‖Yale University, Department of Emergency Medicine, New Haven, Connecticut
#University of Michigan Medical School, Department of Medical Education, Ann Arbor, Michigan

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted April 30, 2015; Revision received September 29, 2015; Accepted September 29, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27270

Introduction: Emergency medicine (EM) milestones are used to assess residents' progress. While some milestone validity evidence exists, there is a lack of standardized tools available to reliably assess residents. Inherent to this is a concern that we may not be truly measuring what we intend to assess. The purpose of this study was to design a direct observation milestone assessment instrument supported by validity and reliability evidence. In addition, such a tool would further lend validity evidence to the EM milestones by demonstrating their accurate measurement.

Methods: This was a multi-center, prospective, observational validity study conducted at eight institutions. The Critical Care Direct Observation Tool (CDOT) was created to assess EM residents during resuscitations. This tool was designed using a modified Delphi method focused on content, response process, and internal structure validity. Paying special attention to content validity, the CDOT was developed by an expert panel, maintaining the use of the EM milestone wording. We built response process and internal consistency by piloting and revising the instrument. Raters were faculty who routinely assess residents on the milestones. A brief training video on utilization of the instrument was completed by all. Raters used the CDOT to assess simulated videos of three residents at different stages of training in a critical care scenario. We measured reliability using Fleiss' kappa and interclass correlations.

Results: Two versions of the CDOT were used: one used the milestone levels as global rating scales with anchors, and the second reflected a current trend of a checklist response system. Although the raters who used the CDOT routinely rate residents in their practice, they did not score the residents' performances in the videos comparably, which led to poor reliability. The Fleiss' kappa of each of the items measured on both versions of the CDOT was near zero.

Conclusion: The validity and reliability of the current EM milestone assessment tools have yet to be determined. This study is a rigorous attempt to collect validity evidence in the development of a direct observation assessment instrument. However, despite strict attention to validity evidence, inter-rater reliability was low. The potential sources of reducible variance include rater- and instrument-based error. Based on this study, there may be concerns for the reliability of other EM milestone assessment tools that are currently in use. [West J Emerg Med. 2015;16(6):871–876.]

INTRODUCTION

As the next phase of competency-based assessment, the Accreditation Council for Graduate Medical Education (ACGME) developed milestones through expert consensus and comprehensive literature reviews.[1,2] The milestones are specialty-specific outcome-based expectations used to evaluate physicians' progress during residency and readiness to complete training, as well as to evaluate residency programs.[3] The ACGME requires semiannual evaluations on 23 emergency medicine (EM) sub-competencies and over 120 milestones. This competency-based assessment makes fundamental skills explicit for learners and allows attendings to evaluate learners based on specific criteria.

The EM milestones were developed with attention to content validity by using an expert panel, building upon the "EM Model," and the work of the American Board of EM.[4] The workgroup that developed the milestones acknowledged that "the next challenge to each residency and the specialty as a whole is the development of objective measures of milestone subcompetency assessment" and the "issues of assessment tool validity and of inter-rater reliability will need to be studied and addressed as various assessment tools are developed, piloted, and put into widespread use."[4]

We must ensure that we are able to accurately measure what we intend to assess; however, as of yet, there are no reliable assessment tools with clear validity evidence. A research group convened to develop such a tool. Workplace-based direct observation is key to assessing performance, and in the ED an attending physician is routinely present to supervise the initial resuscitation of a critically ill patient. The purpose of this study was to design and validate a tool, the Critical Care Direct Observation Tool (CDOT), which allows for direct observation of EM residents on multiple ACGME milestones during the first several minutes of a critical resuscitation.

METHODS

Study Design and Setting

A multi-center, prospective, observational study was conducted at eight academic institutions distributed throughout the country. The study was approved by the review boards at each participating site and was deemed to be exempt as an educational tool without identification of human subjects. This study was part of the Medical Education Research Certification (MERC) program and the Council of Residency Directors (CORD). The research team was composed of five faculty members, two fellows, and one resident from U.S. academic centers.

Instrument Development

The CDOT's design was based on the ACGME EM Milestones.[2] The researchers met via in-person meetings and through monthly conference calls to create the tool using a modified Delphi process.[5] Consensus was reached that an ideal tool would 1) evaluate multiple milestones in an efficient manner using direct observation, 2) be easy to use and be generalizable, and 3) include reliability and validity evidence. When discussing what clinical scenarios would be optimal for direct observation, the panel determined that resuscitations of unstable patients often requires the resident to demonstrate breadth of knowledge, advanced patient care, team management, and communication skills. Further, faculty are routinely present during resuscitation; hence, assessment would not require additional observation time.

The CDOT was developed and revised in spring of 2013 following the principles of content validity, response process, and internal structure (Table).[6] Evidence of content validity is found in the table, with the use of language from the milestones in order to avoid ambiguity and to improve individual test item quality. We added clarification to several items to align with direct observation assessments.

Validity evidence supporting internal structure and response process are noted in the table. Further, the scoring algorithm was derived directly from the ACGME milestones. We modified the scoring categories to decrease ambiguity and consequentially minimize variability in the final score.[9] Feasibility and response process were determined by field testing and revision of the instrument.

Study Protocol

Part one of the study focused on building validity evidence for the CDOT (Table). After the initial CDOT was designed, ensuring content validity for the purpose of response process validity, each of the physician investigators piloted the tool on 29 resident field observations. Feedback was solicited on the performance of the CDOT, and the tool was subsequently revised. Two versions of the CDOT were developed: a Checklist approach (Appendix 1) and a Milestone Rating Scale (Appendix 2), assessing nine of the 23 sub-competencies.

Part two of the study involved evaluating the two versions of the CDOT for inter-rater reliability (internal structure). Faculty reviewed a standardized video of three residents with different levels of training caring for a patient with an aortic dissection and scored them with both CDOTs. The training provided to the faculty assessors consisted of a 10-minute training video introducing the elements of the tool prior to the three resident-patient encounters but did not give specific instruction on how to implement the CDOT. The Checklist format was used with the video review a total of 25 times, and the sub-competency Milestone level CDOT was used a total of 16 times.

Data Analysis

We coded Checklist responses as categorical variables: "Not Applicable," "Not Performed," "Performed Partially," and "Performed Adequately." For each checklist item, inter-rater reliability was estimated using Fleiss' kappa.[7] We compared values of kappa to Landis and Koch's levels of

Table. Three major soures of test validity.[6]

	Definition	Validity evidence for instrument
Content	The extent to which test content and the construct of interest are matched. Evidence of content validity may include test blueprint to match content to construct, the use of experts in the field, literature and guidelines (e.g., milestones) to determine content match with construct.	1) Using language from the milestones, 2) Involving an expert panel of EM residency leaders from six academic institutions, 3) Using a modified Delphi approach, and 4) Utilizing an assessment blueprint based on a review of each of the EM ACGME sub-competencies and determining the appropriateness of each for incorporation into the direct assessment tool
Response process	The cognitive and physical processes required by the assessment also represent the construct. Decisions for response process validity include: the choice for global score versus checklist; analysis of individual responses; debriefing of respondents; and quality assurance and control of assessment data.	1) Explicit scoring algorithms directly related to the underlying construct, 2) By the judgments of the experts regarding the scoring, 3) Adjustment of scoring responses, 4) Field testing and revision
Internal structure	Assessment content and processes provide data about learner performance relevant to the construct. Internal process refers to how assessment transforms the data into a score that represents the construct. Evidence of internal structure includes: statistical characteristics of items and option functions; factor analysis.	*Reliability of reproducibility of scores *Inter-item correlations

EM, emergency medicine; *ACGME*, Accreditation Council for Graduate Medical Education

inter-rater agreement.[8]

For the Milestone Rating Scale, respondents scored residents using the standard milestone levels. The "NA" and zero ratings were treated as missing data. We estimated inter-rater reliability using the intra-class correlation type 1 (ICC1), which estimates the percentage of the rating variance attributable to differences in trainee performance level. Higher ICC1 scores indicate more inter-rater agreement with ICC1=0.80 indicating adequate agreement. We performed all analyses in R version 2.11.1 (R Development Core Team, 2010) using the "irr" package version 0.83 for the computation of Fleiss' kappa and the "multilevel" package version 2.3 for the computation of ICC1.

RESULTS

When the videos were scored, all possible responses from the checklist categories were used. The tool utilization demonstrated adequate response process; however, both the Checklist format and Milestone Rating Scale format were found to have very poor inter-rater reliability. In other words, the faculty could not reliably determine the score of each resident despite all viewing the same performance. Fleiss' kappa of each of the 19 items measured on the Checklist format of the CDOT was near zero for most items and categorized as "slight agreement" for only one item (Figure 1). There was pronounced variability in the raters' use of the "not applicable" category on the Checklist CDOT format; thus, we also calculated Fleiss' kappa excluding this category. However, they were found to be just as low—ranging from -0.04 to 0.25.

The Milestone Rating Scale CDOT had a total of nine items. Mean ratings for each item were low, ranging from 2.26 to 2.83 with an acceptable amount of variability (SD ranging from 0.66 to 1.16), and raters used the full range of the scale for most items.

Rater agreement reliability was near zero (Figure 2). The Intra-class correlations (ICC1) statistics were near zero for all items except one (PC4), which had an estimated reliability of 0.13 (see Figure 2). None of these ICC1 statistics approached the acceptable level of inter-rater agreement of ICC1=0.80. This is due largely to the wide range in ratings for each trainee on each item. Each trainee was rated both low (1 or 2) and high (3, 4 or 5) by at least one rater on every item, and trainee mean ratings did not differ by more than one rating point for any item. This pattern was not due to raters' "hawk/dove" differences. After adjusting ratings so each rater's mean rating was zero, ICC1s remained low (from -0.04 to 0.019).

DISCUSSION

In order to adequately implement milestones, educators need objective, reliable assessment tools with data to support the tool's validity. Although the CDOT was designed with attention to sources of validity evidence, this study found disconcerting results. While each milestone was used

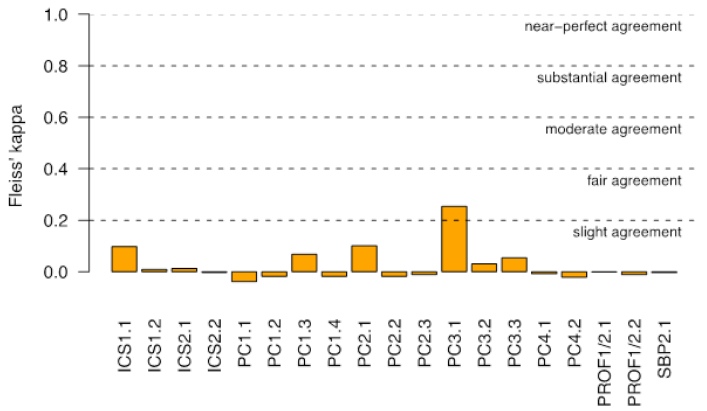


Figure 1. Fleiss kappa for checklist critical care direct observation tool items.

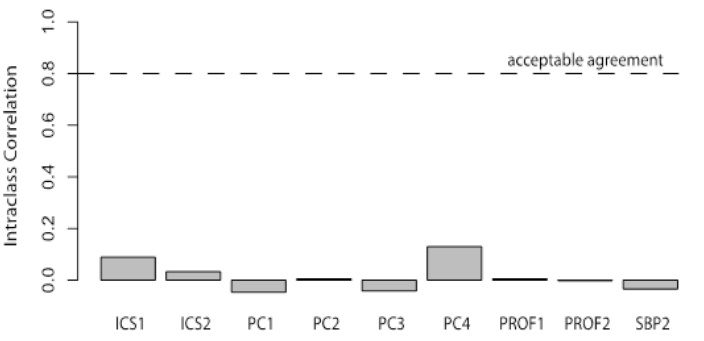


Figure 2. Intra-class correlations for sub-competency milestone level CDOT.
CDOT, Critical Care Direct Observation Tool

essentially verbatim, when used to assess standardized video performance neither the Milestone Checklist nor the Milestone Rating Scale CDOT demonstrated reliability. The reliability analysis demonstrates that essentially raters could not agree on the appropriate scores of the three residents. This is particularly troubling because similar instruments have already been widely adopted by EM programs for resident assessment. Based on the data, the authors do not believe the inter-rater disagreement for both instruments was due to inadequate scale use, range restriction effects, rater bias (hawk/dove), or instrument format.

The underlying issue is error; namely, variance that is not explained by the model. There are two domains of potentially reducible error to consider—rater error and instrument error. This study used the assumption that faculty routinely assess residents on the EM milestones and, for this reason, volunteer faculty raters using direct observation could accurately and consistently judge the performance of videotaped residents. But this clearly did not happen. When analyzing the potential sources of rater error, the following types of bias may be playing a role: rater inconsistency, severity and leniency, frame of reference, central tendency, and the halo effect. Workplace assessments in medicine require judgment on the part of the rater, which suggests that there may be no such thing as a purely objective interpretation of assessment results.[18]

Rater inconsistency in this context occurs when a faculty member fails to apply the rating scale in the same manner as other faculty members. This diminishes the tool's ability to differentiate between higher achieving and lower achieving residents.[10] Second, leniency and severity biases undermine inter-rater reliability. Leniency bias occurs when a faculty member gives high scores even when a performance is not deserving of such as score ("doves") and severity bias occurs when raters give low scores despite good performance ("hawks").[10-12] Third, faculty may tend to use their own clinical practice style as a frame of reference for clinical assessment rather than adhering to the agreed-upon standard.[13] Further, faculty are experts and may take shortcuts in patient care due to expert intuitive judgment that can impact scoring. Fourth, faculty tend to avoid extreme positions on a rating scale, resulting in a central tendency on their assessments, essentially restricting the intended range of the rating scale.[10] Lastly, when faculty personally know or identify with a particular learner, there may be a tendency to assess that learner more positively (the halo effect).[10]

Intuitively, rater training should minimize these types of rater error. Training assessors how to use assessment tools has been shown to improve reliability in some studies,[14] while having little effect in others.[15] The clinical experience of faculty assessors as well as their knowledge of the intricacies of milestones may affect their trainability.[16] Regardless, the common practice at most residencies is to have faculty without training assess residents on the milestones in the clinical setting.

The second domain of error, instrument error, refers to the difference between the actual skill and the value measured by the rating instrument. When it comes to rating scales, some advocate for making the tool appear less subjective by removing scales (1-5) and inserting yes/no type checklists. In this study, there was no difference between using a checklist compared to a scale. The literature does not definitively advocate for the use of one rating scale over another.[12] Using the EM milestones themselves within the CDOT tool may be another source of instrument error. While the use of milestones phrasing improved content validity, the wording of many of the milestones is relatively broad and may not be useful when used to rate specific behaviors.[17] One final source of instrument error may be that the number of items is too large, resulting in cognitive overload. Raters may not be able to keep in their working memory all of the items on the scale, resulting in an incorrect score. Unfortunately, there is no true measurement or standard to evaluate how accurately we are judging our residents.

Beeson et al. recently published a study demonstrating that the internal consistency of the subcompetencies using a Cronbach's alpha coefficient was high, although the study did

not specifically look at workplace-based assessment and rater variability. This study used the milestone scoring for each resident that was submitted to ACGME. The accompanying editorial noted "there is a need for additional validity evidence from multiple sources, evaluation of potential limiting bias, and defining of the appropriate role of milestones in assessment.

Limitations

Our study had several limitations. The definition of critical care resuscitation likely varies widely among EM physicians; thus, the context of the case and the idiosyncrasy of faculty judgments are limitations. Faculty training was another limitation. While there was a brief video on rater training, specific attention to helping faculty understand the milestones may have improved the inter-rater reliability. However, modeling after current practice, faculty are routinely rating residents with minimal training. Additionally, the use of videos to collect validity evidence may not represent the actual functioning of the tool in a clinical setting. Finally, the clinical experience of our assessors was not documented. It is possible that the lack of inter-rater reliability could be attributed to a difference in the assessment inferences used by faculty based on the faculty members' experience.

Future Directions

What is to be done? First, we continue to advocate for direct observation of workplace-based assessment as a component of milestone evaluation. The use of milestones provides a framework for the very important conversation between faculty and trainee to describe performance and identify areas of excellence and those areas needing improvement. While the CDOT instrument is limited in reliability, it may be effective as a tool to use as a framework for discussion during direct observations of critical patients.

When there are reliability and validity issues, Van der Vleuten argues for the use of programmatic assessment using multiple modalities and lower stakes assessment to achieve a more complete picture of the learner.[18] It is imperative to understand that the milestones are not assessment tools themselves but are constructs against which we reference resident performance. As final milestone assessment for EM residents is a high-stakes summative assessment, the goal of the Clinical Competency Committee should be to incorporate various assessment tools from multiple individuals regarding resident performance in order to make assessments that are as reliable and valid as possible.

CONCLUSION

Although EM residents are currently being assessed on milestones, the validity and reliability of tools for such assessment have yet to be determined. Implementing milestones-based evaluation is a formidable challenge as we must generate evidence to inform the development of assessment tools. This study was a rigorous attempt to collect validity evidence for an EM milestone direct observation instrument. Despite nearly verbatim use of the EM milestones during construction of this tool, while maintaining content validity, the resulting responses were not reliable and were fraught with variability. This may be secondary to rater- and instrument-based error. However, based on this study, there are significant concerns for the reliability of other EM milestone assessment tools in use that have not been examined in terms of their reliability and validity.

Address for Correspondence: Meghan Schott, MD, Tacoma Emergency Care Physicians, PO Box 5215, Tacoma, WA 98415. Email: meghan.schott@tecp.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Swing SR. Assessing the ACGME General Competencies: General Considerations and Assessment Methods. *Acad Emerg Med.* 2002;9(11):1278-88.
2. Nasca TJ. The next step in the outcomes-based accreditation project. *ACGME Bulletin.* May 2008.
3. Nasca TJ, Philibert I, Brigham T, et al. The Next GME Accreditation System — Rationale and Benefits. *N Engl J Med.* 2012;366(11):1051-6.
4. Beeson MS, Carter WA, Christopher TA, et al. The Development of the Emergency Medicine Milestones. *Acad Emerg Med.* 2013;20(7):724-9.
5. Clayton MJ. Delphi : a technique to harness expert opinion for critical decision-making tasks in education. *Educ Psychol. (Dorchester-on-Thames. Print).* 1997;17(4):373-86.
6. Downing SM and Haladyna TM. Validity and its Threats. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. New York: Taylor & Francis; 2009.
7. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76:378-82.
8. Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977;33(1):159-74.
9. Downing SM and Yudkowsky R. *Assessment in Health Professions Education*. New York: Taylor & Francis; 2009.

10. Iramaneerat C and Yudkowsky R. Rater Errors in a Clinical Skills Assessment of Medical Students. *Evaluation & the Health Professions.* September 1, 2007 2007;30(3):266-83.
11. Marchegiani L, Reggiani T, Rizzolli M. Severity vs. Leniency Bias in Performance Appraisal: Experimental Evidence. *Bozen Economics & Management Paper Series.* 2013.
12. Pangaro LN and Holmboe ES. Evaluation Forms and Global Rating Scales. In: Holmboe ES, Hawkins RE, eds. *Practical Guide to the Evaluation of Clinical Competence*. Philadelphia: Mosby; 2008.
13. Kogan JR, Hess BJ, Conforti LN, et al. What Drives Faculty Ratings of Residents' Clinical Skills? The Impact of Faculty's Own Clinical Skills. *Acad Med.* 2010;85(10):S25-S28.
14. Feldman M, Lazzara EH, Vanderbilt AA, et al. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health.* 2012;32(4):279-86.
15. Cook D, Dupras D, Beckman T, et al. Effect of Rater Training on Reliability and Accuracy of Mini-CEX Scores: A Randomized, Controlled Trial. *J Gen Intern Med.* 2009;24(1):74-9.
16. Govaerts MJB, Schuwirth LWT, Vleuten CPM, et al. Workplace-based assessment: effects of rater expertise. *Adv in Health Sci Educ.* 2011;16(2):151-65.
17. Accreditation Council for Graduate Medical Education. ACGME e-Communication - April 28, 2014.
18. Van der Vleuten C and Schuwirth L. Assessing professional competence: from methods to programmes. *Med Educ.* 2005;39(3):309-17.
19. Beeson MS, Homboe ES, Korte RC et al. Initial Validity Analysis of the Emergency Medicine Milestones. *Acad Emerg Med* 2015;22:838–44.
20. Love JN, Yarris LM, Ankel FK. Emergency Medicine Milestones: The Next Step. *Academic Emergency Medicine* 2015;22:847-8.

Development and Implementation of an Emergency Medicine Podcast for Medical Students: EMIGcast

Andrew Lichtenheld, BS
Mari Nomura, BA
Nicholas Chapin, BA
Trenton Burgess, BS
Joshua Kornegay, MD

Oregon Health & Science University, Department of Emergency Medicine, Portland, Oregon

Section Editor: David Wald, DO
Submission history: Submitted April 30, 2015; Revision received September 1, 2015; Accepted September 26, 2015
Electronically published October 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27293
[West J Emerg Med. 2015;16(6):877–878.]

INTRODUCTION

Podcasts, episodic digital audio recordings downloaded through web syndication or streamed online, have been shown to be an effective instructional method in undergraduate health professions education, and are increasingly used for self-directed learning.[1-6] Emergency medicine (EM) has embraced podcasting: over 80% of EM residents report listening to podcasts and a substantial number identify podcasts as the most valuable use of their educational time.[4] Despite proven efficacy in undergraduate medical education and remarkable popularity with EM residents and attendings, there remain few EM podcasts targeted to medical students.[5] Given that podcast effectiveness correlates with how well content matches the listener needs, a podcast specific to EM-bound medical students may optimally engage this target audience.[6]

OBJECTIVES

Our educational goals involved both content and process: 1) to produce a podcast delivering core EM content targeting specific needs of medical students interested in EM; and 2) to provide EM-bound medical students an opportunity to develop the ability to use podcasts as a self-directed educational modality. We measured time, costs, and resources as feasibility data and tracked podcast uptake as an initial measure of learner acceptability.

CURRICULAR DESIGN

We developed this educational advance based upon established instruction design methods:

Problem Identification, General and Targeted Needs Assessment

Our student EM interest group (EMIG) identified a gap in podcasts meeting learning needs of EM-bound medical students through peer discussions. A general needs assessment included a literature search and review of existing podcasts for medical students and with EM content, and informed a targeted needs assessment comprised of discussions with students and physicians regarding students' needs and interests.

Goals and Objectives

We designed the curriculum to prepare medical students to:

1. Demonstrate the ability to access podcasts and incorporate them into strategies for self-directed learning.
2. Develop familiarity with important themes and terminology in clinical EM practice.
3. Analyze and summarize discussions of core EM content in order to apply it to clinical practice.
4. Specific learner objectives were also developed for individual podcast.

Educational Strategies

We chose the podcast modality for its effectiveness, popularity and accessibility.[4] Under EM faculty guidance, a group of four medical students identified EM topics of importance to their peers. Content categories included: clinical conditions encountered in EM; logistics of applying to residencies, training, and working in a career in EM; and ethics encountered in EM. After defining a content outline, targeted learner objectives were developed. Medical students then identified content experts consisting of EM residents, nurses, and faculty and invited them to participate in student-led interviews. For each podcast an interview protocol was created with an anticipated discussion flow structured to elicit content matching learner objectives.

Implementation, Resources and Logistics

Portable audio recording equipment and an online blog and podcasting platform were purchased with funds obtained from an educational grant. Startup equipment and costs totaled $400. We branded our podcast "EMIGcast," reflecting the involvement of EMIG. Five episodes were developed, recorded, edited for clarity and brevity (average final length 33 minutes), and web-syndicated through the EMIGcast website and iTunes over a five-month period.[7] Software with the podcasting platform was used to capture feasibility data including downloads of the audio content.

IMPACT/ EFFECTIVENESS

We present the development and implementation of a podcast specifically designed to achieve content and process goals targeting EM-bound medical students. We found the podcast feasible to implement and acceptable to our learners. Data collected from the first 20 weeks of the "EMIGcast" podcast demonstrates an average of 148.5 downloads per month (698 total downloads) with a consistent increase in monthly downloads (Figure). While the absolute number of downloads is modest, the upward trend suggests growing acceptability and the number of downloads represents a scalability exceeding what we have previously achieved with EMIG lectures and panel discussions. We have also seen that episodes continue to be downloaded months after release, implying that the asynchronous and longitudinal availability of the podcast may be valuable to students over time. The student-initiated format may help with learner buy-in and the affiliation with EMIG provides name recognition and an inherent benefit for sustainability. We plan to identify new student leaders in incoming EMIG members each year and use consistent longitudinal faculty leadership to maintain institutional history and adherence to the overarching goals of the intervention. Future planned evaluations of the podcast include survey of listeners to better measure success in achieving learning objectives and to better characterize the educational impact of this educational intervention.

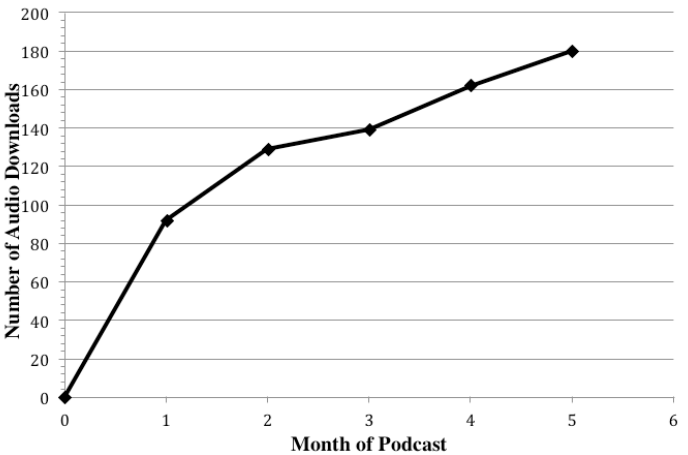


Figure. Total downloads by month of "EMIGcast" syndication

Address for Correspondence: Josh Kornegay, MD, Oregon Health & Science University, 3181 SW Sam Jackson Park Rd. CDRC - West, Portland, OR 97239. Email:kornegaj@ohsu.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Vogt M, Schaffner B, Ribar A, et al. The impact of podcasting on the learning and satisfaction of undergraduate nursing students. *Nurse Educ Pract*. 2010;10(1):38-42.
2. Kalludi SN, Punja D, Pai KM, et al. Efficacy and perceived utility of podcasts as a supplementary teaching aid among first-year dental students. *Australas Med J*. 2013;6(9):450-7.
3. White JS, Sharma N, Boora P. Surgery 101: evaluating the use of podcasting in a general surgery clerkship. *Med Teach*. 2011;33(11):941-3.
4. Mallin M, Schlein S, Doctor S, et al. A survey of the current utilization of asynchronous education among emergency medicine residents in the United States. *Acad Med*. 2014;89(4):598-601.
5. Berger E. Web 2.0 in emergency medicine; specialty embracing the future of medical communication. *Ann Emerg Med*. 2012;59(4):A21-3.
6. Matava CT, Rosen D, Siu E, et al. eLearning among Canadian anesthesia residents: a survey of podcast use and content needs. *BMC Med Educ*. 2013;13:59.
7. EMIGcast website. Available at: http://www.emigcast.com/.

Ready for Discharge? A Survey of Discharge Transition-of-Care Education and Evaluation in Emergency Medicine Residency Programs

Fiona E. Gallahue, MD*
Amy E. Betz, MD*
Jeffrey Druck, MD†
Jonathan S. Jones, MD‡
Boyd Burns, DO§
Gene Hern, MD, MS¶

*The University of Washington, Division of Emergency Medicine, Seattle, Washington
†The University of Colorado, Department of Emergency Medicine, Aurora, Colorado
‡The University of Mississippi Medical Center, Department of Emergency Medicine, Jackson, Mississippi
§The University of Oklahoma School of Community Medicine,Tulsa, Oklahoma
¶Alameda Hospital System-Highland Hospital, Department of Emergency Medicine, Oakland, California

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted April 30, 2015; Revision received September 4, 2015; Accepted September 16, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27298

This study aimed to assess current education and practices of emergency medicine (EM) residents as perceived by EM program directors to determine if there are deficits in resident discharge handoff training. This survey study was guided by the Kern model for medical curriculum development. A six-member Council of EM Residency Directors (CORD) Transitions of Care task force of EM physicians performed these steps and constructed a survey. The survey was distributed to program residency directors via the CORD listserve and/or direct contact. There were 119 responses to the survey, which were collected using an online survey tool. Over 71% of the 167 American College of Graduate Medical Education (ACGME) accredited EM residency programs were represented. Of those responding, 42.9% of programs reported formal training regarding discharges during initial orientation and 5.9% reported structured curriculum outside of orientation. A majority (73.9%) of programs reported that EM residents were not routinely evaluated on their discharge proficiency. Despite the ACGME requirements requiring formal handoff curriculum and evaluation, many programs do not provide formal curriculum on the discharge transition of care or evaluate EM residents on their discharge proficiency. [West J Emerg Med. 2015;16(6):879–884.]

Millions of patients are seen in the emergency department (ED) with approximately 86% rate of discharge.[1] The discharge transition of care is the most commonly performed handoff in the ED and yet, most studies have focused on the handoffs between providers. Discharge is the handoff from provider responsibility to patient responsibility for care. This is a complex process representing a time of significant vulnerability for patients. Safe and effective transfer of responsibility for a patient's medical care relies on effective provider communication with patient comprehension of discharge instructions. Studies have demonstrated that patients discharged from the ED have significant gaps in their understanding of this information.[2,3] There is some literature to suggest that the quality of verbal communication at discharge by emergency providers is incomplete and leaves little room for patients to ask questions.[4] There is evidence to suggest that ineffective communication between providers and patients is a source of error in the discharge period.[5,6] It has been demonstrated that residents overestimate the effectiveness of their communication.[7] Additionally, residents may not recognize patient factors that place patients at high risk for readmission.[8]

Evidence on how to ensure ideal handoffs has been limited, but multiple sources have identified process standardization as an opportunity for quality

improvement.[9,10] Accordingly, standardization of handoffs was made a National Patient Safety Goal by the Joint Commission in 2006.[11] The American College of Graduate Medical Education (ACGME) has identified education and evaluation of care transitions as an educational mandate in training programs. The ACGME states "formal educational activities that create a shared mental model with regard to care transitions are necessary" and that "evaluation through direct observation of residents/fellows by faculty members is required to ensure residents'/fellows' abilities to perform standardized, effective, efficient handoffs."[12]

Although there are clear mandates to ensure that handoffs are standardized, and formal handoff curriculum and evaluation are provided to emergency medicine (EM) residents, there is no information available to identify the current practices of EM training programs on the discharge transition of care.

The objectives in this study were to (1) assess the current scope of discharge training among EM residency programs by surveying their residency leadership, (2) assess current educational and evaluation practices related to discharge training, and (3) identify whether additional training is necessary based on the current practices and perceived competencies.

METHODS

This survey study was guided by the six-step Kern model for medical curriculum development.[13] A similar study by the CORD Transitions of Care taskforce membership was performed on general handoff training, which provided a foundation for this study.[14] The first two stages–problem identification and creation of a targeted needs assessment– were the goals of this study. The latter four stages of the Kern model include determining the goals and objectives of the curriculum, developing educational strategies for teaching the curriculum, implementing the curriculum, and receiving feedback on and evaluating the curriculum. The application of this model provides the opportunity to eventually develop a discharge handoff curriculum for EM residents. Establishing validity evidence was an important consideration throughout the process; validity evidence comes in the form of content, response process, internal structure, relationship to other variables, and consequences.[15] We conducted a review of discharge literature to survey common practices and sources of error and to discover EM discharge education techniques. One identified problem was the lack of formal emergency discharge education available or required during residency training. The survey went through a thorough development process using the expertise of those involved in its development to contribute to the validity of its content.[15] This process was an iterative approach by the authors who are on the Council of Residency Directors in EM (CORD) Transitions of Care Committee (FG, JD, JJ, BB, HGH) or have served as a director of Quality Assurance and Improvement (AB). The initial survey was developed by two authors (HGH, JD) based on focus group comments and suggestions from membership during a CORD Transitions of Care Committee meeting (approximately 20 faculty in attendance). We sent the draft survey to the authorship team via SurveyMonkey® (an online survey development cloud-based service) to complete and suggest further content changes. This was repeated twice more until the team felt that the final questionnaire best addressed the areas where knowledge content gaps were identified by the focus group. The survey focused on elucidating current practices of discharge training and clinical practices within the clinical environments, including current education offered, perceptions of best educational practices, methods of resident evaluation and perceived competence of residents. Multiple-choice questions were the primary vehicle for the response process.

The institutional review board at Alameda County Medical Center (Highland Hospital, Oakland, CA) granted exempt approval for this study. Members of CORD were invited to complete the survey electronically. The CORD e-mail listserve is exclusive to educators in EM residency programs and includes associate, assistant and primary residency program directors from the 167 ACGME-accredited EM residency programs. Program leaders were recruited via the CORD listserve from March to April 2014. The survey was opened for six weeks in which 87 identified programs responded. Duplicate responses were reviewed and clarified with program directors directly in April 2015 (22 programs). Two programs requested their duplicate responses be deleted and completed new surveys to accurately represent their current practice. All other programs selected their most accurate survey responses. Direct emails were then sent to program directors with links to the survey in April 2015; 32 additional programs completed the survey and the survey was closed on April 23, 2015. Only programs that identified their program name were included in the study to ensure that all surveyed programs represented ACGME-accredited programs and to avoid potential duplicative responses.

RESULTS

A total of 119 programs were surveyed, making the overall response rate 71.2% of the 167 currently accredited ACGME EM residency training programs. A majority of programs indicated that residents are given informal education regarding discharge training by senior residents and faculty (87.4%); just under half provide formal curriculum at orientation (42.9%) and/or outside of orientation (5.9%). A small percentage of programs offer no training (6.7%) (Table 1). Over half of programs felt that optimal discharge training would be formal curriculum offered at orientation (76, 63.9%) and/or outside of orientation (63, 52.9%).

Most residency programs reported using a structured discharge system in the ED (100, 84%%), while a small minority report using none (13, 10.9%) or being unsure if

Table 1. Transition-of-care survey results representing 119 respondents.
Q1: Types of discharge training offered to residents/students (more than one response acceptable).
Q2: What training around discharge processes do you believe would be best to provide to residents/students (more than one response acceptable).

Answer options	Percent response (count): Q1	Percent response (count): Q2
No training	6.7% (8)	1.7% (2)
Specific formal training regarding discharges during initial orientation to the program	42.9% (51)	63.9% (76)
Structured workshops/classes to teach proper discharge processes during residency (not in orientation)	5.9% (7)	52.9% (63)
Instruction by attending/senior resident within the clinical environment	87.4% (104)	69.7% (83)
Distributed educational packets/guides	9.2% (11)	18.5% (22)
Formal evaluation of residents on competency in performing effective discharges	14.3% (17)	58% (69)
Informal evaluation of residents on competency in performing effective discharges	39.5% (47)	24.4% (29)
Other	N/A (3)*	N/A (1)**

*3 responses: 1) Discuss during Morbidity & Mortality conference. 2) Formal evaluation of effective discharges is covered loosely in chart reviews, call backs are performed intern year. 3) Grand rounds presentation of ED Discharges.
**1 response: 1) I don't know.

this is provided in their main ED (4, 7.6%). For those with a structured process of discharge most reported that this is being performed most of the time (35, 35%) or always (56, 56%). A majority of programs reported providing structured written modifiable written discharge instructions (108, 90.8%) and bidirectional conversations with patients (91, 76.5%) (Table 2**).** A majority of training programs reported that their current process is safe and effective or extremely safe and effective (78, 66%) but a significant number reported their process to be somewhat safe and effective (41, 34%). Most programs do not formally evaluate their residents for discharge proficiency (88, 73.9%). Of those that reported resident evaluation of discharge proficiency is routinely performed (more than one response was acceptable), 25 programs (21%) report the evaluation is completed as part of required direct observations or other activities, nine (7.6%) reported it is completed through written feedback/evaluation of performance on ED rotations, and two programs (1.7%) reported a formal assessment of discharge proficiency is completed on junior residents as part of a checklist or similarly structured evaluation.

All programs reported a variety of tools to assist in the discharge process with only 34, or 43%, of the respondents being satisfied or extremely satisfied with these tools (Table 3). Over two-thirds of programs reported that key elements of discharge conversations, such as diagnosis, education, prescriptions, follow up, return precautions or assessment of understanding, are documented in the physician note (80, 67.2%). This information is not routinely documented in 31 programs (26.1%), and eight respondents were not sure if this is included in their documentation (6.7%).

Over three-quarters (90, 75.6%) of program leadership reported that junior level residents (equal to or less than eight months in the ED) are "somewhat competent" in their discharge competency. Almost a quarter felt that their junior residents were "competent" (28, 23.5%). One program (0.8%) reported their junior residents were extremely competent (1, 0.8%). For senior level residents, described as residents with over eight months experience, their program leadership identified them as competent (83, 68.7%) or extremely competent (26, 21.8%). A minority of programs reported their senior level residents as "somewhat competent" in their discharge skills (10, 8.4%). None of the respondents reported their junior or senior level residents to be incompetent in their discharge abilities.

DISCUSSION

These results provide insight into the discharge educational practices and clinical training experience surrounding discharge of EM residents as reported by their program leadership. Standardized formal training and evaluation is not the current norm at most programs. Most formal training provided to EM residents is at their orientation, and few programs offer formal educational opportunities beyond the first few weeks of training. Since a majority of program leaders indicated that ideal educational practices would include formal training at orientation and/or workshops or classes outside of orientation, programs may value structured education of discharge competencies but may be constrained by other limitations such as faculty time, didactic scheduling or curriculum availability. The same gap was seen between current training and ideal training for evaluation processes. While most programs do not perform formal evaluations on their residents' discharge competency, a majority of programs identify that this would be an ideal

Table 2. Results represent 119 respondents.
Q: Which of the following is included in your standard discharge process? More than one response is acceptable.

Answer options	Percent response (count)
Provision of any prewritten non-modifiable instructions	30.3% (36)
Provision of any structured written instructions that allow for modification	90.8% (108)
Physician routinely has bidirectional conversation with patient regarding diagnosis, education, prescriptions, follow up and reasons to return to the ED	76.5% (91)
Teach back method (or similar) routinely employed to assess patient understanding of their diagnosis, education, prescriptions, follow up and reasons to return to the ED	10.9% (13)
Final discharge routinely completed by nursing	68.9% (82)
Final discharge routinely completed by physician	10.9% (13)
Final discharge routinely completed by either nursing or physician	22.7% (27)
Other	N/A (2*)

ED, emergency department
*2 responses: 1) Provide follow up physician or clinic. 2) Nursing employs teach back with patients.

practice. This implies that EM program leadership values formal evaluation of their residents' discharge competency but they may be constrained by limitations such as a recognized evaluation tool and/or faculty time. These data also suggest that while program leadership values discharge competency training and evaluations, this education may not be valued as a high priority since most programs perceive their senior level residents to be competent. Although it is difficult to fully endorse competency without a standardized evaluation process, there is support that informal evaluation may be valid in identifying residents' clinical competencies.[16]

LIMITATIONS

The fact that this study relies on perceptions of program leadership is a major limitation as there is no gold standard to formally measure discharge competency even for programs providing formal evaluation of their residents' discharge competency. While most program leadership feel that their current process of discharge is "safe and effective," this perception may be limited. Given that each respondent based their program results on his or her perceptions of the discharge education and performance of residents within the clinical environment, construct underrepresentation and irrelevant variance represent threats to the validity of clinical performance ratings in this study.[17] Program leadership could have responded based on too few or incomplete observations of residents' clinical behavior or responded with low-reliability ratings. Survey, rater and recall bias could have affected these results.

Program leadership relies on general gestalt that their residents are competent in their discharge proficiency since most are not evaluating this competency. Program leadership may not identify any limitations in senior resident discharge competency because there may be larger, more systematic failures of discharge communication with patients and/or caregivers. In this scenario, resident performance of discharge may be at an acceptable level at the departmental level but departmental expectations of patient discharge competency may not be meeting the patient needs to create safe and effective discharges. Standardizing the process of handoffs between providers in the hospital environment has demonstrated improvement in the quality of communication and increased patient safety in the clinical arena.[9] Current discharge literature suggests that there may be similar communication improvements to be made in EM around discharge.[3-5] That over one-third of programs report their current discharge process is only "somewhat safe and effective" suggest that there may be quality gaps in the current institutionally acceptable patient discharge processes.

The conflation of the concept of "safety" and "effectiveness" may be another limitation of this survey. Programs were asked about safe and effective discharges in a single question. These two concepts may be inappropriately linked together and it may be that safety may exist without being effective and vice versa. This may represent a construct error in the survey design and affected results of this survey.

Lastly, a major limitation of this study was the gap of approximately one year between the surveys of the first cohort of surveyed programs (87) and the second (32). Although it is unlikely that most programs changed their educational practices dramatically within that time period, it is possible.

Further work should focus on the more structured assessment of resident discharge competency through direct observation and evaluation of resident performance to corroborate program leaders' assessment of resident competence. These evaluation tools should then be validated and studied in a clinical setting with specific process measures and patient outcomes. Following Kern's six-step model for curriculum development, the next step would be to create specific goals and objectives of the discharge curriculum and develop educational interventions aligned with these goals. Curricular tool suggestions that have been made to structure

Table 3. Q: Do you use any of the following tools to assist in the discharge process? More than one response is acceptable.

Answer options	Percent response (count)
Automated reminders within the computer interface	42.9% (51)
Written template or other written aids (badge card checklist)	24.4% (29)
Mnemonics	1.7% (2)
Teach back or similar method	3.4% (4)
None	38.7% (46)
Other	2*

*2 responses: 1) Nursing feedback when discharge performed improperly. 2) Pre-populated recommendations from nursing triage such as smoking cessation for smokers, blood pressure re-check for patients with high blood pressure at triage; all patients without a primary care provider are provided a printout of the free and low cost medical, dental and mental health resources in the community.

education and evaluation for provider-to-provider handoffs might be used directly or modified to educate residents in the discharge transitions of care.[18]

CONCLUSION

The results of this targeted needs assessment indicate a lack of structured training and assessment of resident discharge competency despite current guidelines for formalized training in all handoffs. Although most programs reported senior residents are competent in discharge proficiency, the residents' training is primarily informal which may lead to significant variability in resident experience and performance. Further research should be aimed at assessing proficiency of resident discharge performance through objective observation with validated evaluation tools. Structured training and assessment recommendations should follow from this research with increased attention to implementing a standard curricular model or toolbox, objective, valid evaluation methods, and identification and management of high-risk discharges.

Address for Correspondence: Fiona E. Gallahue, MD, University of Washington, Harborview Medical Center, 325 Ninth Ave., Division of Emergency Medicine, Box 359702, Seattle, WA 98104. Email: fionagallahue@gmail.com.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. 2011 National Hospital Ambulatory Medical Care Survey. Available at: http://www.cdc.gov/nchs/. Accessed Apr 23, 2015.
2. Agency for Healthcare Research and Quality. Improving the Emergency Department Discharge Process: Environmental Scan Report. Available at: http://www.ahrq.gov/professionals/systems/hospital/edenvironmentalscan/index.html. Accessed Apr 23, 2015.
3. Samuels-Kalow ME, Stack AM, Porter SC. Effective discharge communication in the emergency department. *Ann Emerg Med*. 2012;60(2):152-9.
4. Vashi A and Rhodes KV. "Sign right here and you're good to go": a content analysis of audiotaped emergency department discharge instructions. *Ann Emerg Med*. 2011;57(4):315-22.
5. Kuan WS and Mahadevan M. Emergency unscheduled returns: can we do better? *Singapore Med J*. 2009;50(11):1068-71.
6. Kripalani S, Jackson AT, Schnipper JL, et al. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. *J Hosp Med.* 2007;2(5):314-23.
7. Chang VY, Arora VM, Lev-Ari S, et al. Interns overestimate the effectiveness of their hand-off communication. *Pediatrics*. 2010;125(3):491-6.
8. Powell CK and Kripalani S. Brief report: Resident recognition of low literacy as a risk factor in hospital readmission. *J Gen Intern Med*. 2005;20(11):1042-4
9. Starmer AJ, Spector ND, Srivastava R, et al. Changes in medical errors after implementation of a handoff program. *N Engl J Med*. 2014;371(19):1803-12.
10. Cheung DS, Kelly JJ, Beach C, et al. Improving handoffs in the emergency department. *Ann Emerg Med*. 2010;55(2):171-80.
11. The Joint Commission. Critical Access Hospital and Hospital National Patient Safety Goals. 2006. Available at: http://www.splashcap.com/JCAHO_2006-NPSG-3D. Accessed Apr 23, 2015.
12. Accreditation Council of Graduate Medical Education. Clinical Learning Environment Review. 2014. Available at: https://www.acgme.org/acgmeweb/Portals/0/PDFs/CLER/CLER_Brochure.pdf. Accessed Apr 23, 2015.
13. Kern DE, Thomas PA, Hughes MT. *Curriculum Development for Medical Education: A Six Step Approach.* Baltimore, MD: Johns Hopkins University Press, 2009.
14. Kessler C, Shakeel F, Hern HG, et al. A survey of handoff practices in emergency medicine. *Am J Med Qual*. 2014;29(5):408-14.
15. Downing SM. Validity: on meaningful interpretation of assessment

data, *Med Educ*. 2003;37(9):830-7.

16. Schwind CJ, Williams RG, Boehler ML, et al. Do individual attendings' post-rotation performance ratings detect residents' clinical performance deficiencies? *Acad Med*. 2004;79(5):453-7.
17. Downing SM and Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. 2004;38(3):327-33.
18. Wohlauer MV, Arora VM, Horwitz LI, et al. The patient handoff: a comprehensive curricular blueprint for resident education to improve continuity of care. *Acad Med*. 2012;87(4):411-8.

Combined Versus Detailed Evaluation Components in Medical Student Global Rating Indexes

Kim L. Askew, MD
James C. O'Neill, MD
Brian Hiestand MD, MPH
David E. Manthey, MD

Wake Forest School of Medicine, Department of Emergency Medicine, Winston-Salem, North Carolina

Section Editor: David Wald, MD
Submission history: Submitted April 30, 2015; Revision received September 4, 2015; Accepted September 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27257

Introduction: To determine if there is any correlation between any of the 10 individual components of a global rating index on an emergency medicine (EM) student clerkship evaluation form. If there is correlation, to determine if a weighted average of highly correlated components loses predictive value for the final clerkship grade.

Methods: This study reviewed medical student evaluations collected over two years of a required fourth-year rotation in EM. Evaluation cards, comprised of a detailed 10-part evaluation, were completed after each shift. We used a correlation matrix between evaluation category average scores, using Spearman's rho, to determine if there was any correlation of the grades between any of the 10 items on the evaluation form.

Results: A total of 233 students completed the rotation over the two-year period of the study. There were strong correlations (>0.80) between assessment components of medical knowledge, history taking, physical exam, and differential diagnosis. There were also strong correlations between assessment components of team rapport, patient rapport, and motivation. When these highly correlated were combined to produce a four-component model, linear regression demonstrated similar predictive power in terms of final clerkship grade (R^2=0.71, CI_{95}=0.65–0.77 and R^2=0.69, CI_{95}=0.63–0.76 for the full and reduced models respectively).

Conclusion: This study revealed that several components of the evaluation card had a high degree of correlation. Combining the correlated items, a reduced model containing four items (clinical skills, interpersonal skills, procedural skills, and documentation) was as predictive of the student's clinical grade as the full 10-item evaluation. Clerkship directors should be aware of the performance of their individual global rating scales when assessing medical student performance, especially if attempting to measure greater than four components. [West J Emerg Med. 2015;16(6):885–888.]

INTRODUCTION

Most medical schools employ a combination of multiple choice testing, standardized patients, and direct observation to evaluate their students' performance in terms of knowledge, skills, and attitudes.[1] A staple among evaluations of medical students in clinical rotations is the assessment of a student by a faculty member or resident using a global rating scale (GRS) with varied components, such as knowledge, rapport, procedural skill and documentation quality. Distinct from other clerkships, most EM clerkships require the use of a GRS evaluation card on every shift because students interact with multiple faculty members over the course of a rotation, as opposed to a sustained interaction with one or two individual faculty members.[1] Most clerkships use these cards because students work with different faculty each shift and because of evidence that shift

cards promote immediate and satisfactory feedback.[2,3] Since students do not consistently work with the same attending and residents over a two- to four-week period, clerkship directors rely on the formative feedback provided on shift cards to gauge a student's aggregate clinical performance.

Over the past six decades, the use of these subjective evaluations has found critics and advocates.[4] In order to maximize feedback and quality of evaluations, many institutions have gone to a criterion-based multi-point scale for multiple attributes. However, problems with multi-point scales include increased complexity in completing evaluations, and the ever-present evaluator who "circles down the middle of the scale."[5] Studies from core general surgery clerkships suggest that faculty evaluations of clinical performance limited to three points may be as effective as larger scales in predicting a student's final grades.[4] Also, internal medicine evaluations of students have shown that breaking down grades to three content areas are also predictive of student performance in several types of evaluations.[6]

However, the emergency medicine (EM) medical education literature contains few publications about the most effective methods to evaluate learners in the clinical setting and the performance of global rating scales in the emergency department (ED) setting.[7] Therefore, we conducted this study to evaluate the degree that the evaluations provided redundant (as defined by highly correlated) data between the various components of a 10-point global rating scale. A secondary goal of the study was to see if a reduced global rating scale, based on concatenating scores from highly correlated rating items, would provide a similar evaluation of medical students' overall performance.

Student:			Faculty:				
Shift:			Date:		Encounters:		
Basic Skills	Fail (0)	Marginal (1)	Concern (1.5)	Pass (2)	High Pass (2.5)	Honors (3.0)	N E
Medical Knowledge							
History							
Physical Exam							
Differential / Treatment Plan							
Presentation							
Professionalism							
Health Care Team Rapport							
Motivation							
Procedural Skills							
Documentation							

Figure 1. Shift evaluation card. (Reverse side not pictured contained area for written feedback on student's strengths and areas for improvement).

METHODS

Study Design

We performed a retrospective evaluation on a pre-existing administrative database containing evaluations of medical student performance during their rotation in EM. Institutional review board approved this study with a waiver of consent.

Study Setting and Population

During the study period, the medical school curriculum included a mandatory EM clerkship in the fourth year of medical school. Students completed 15 eight-hour shifts in the ED and generally worked with a single faculty member for the entire shift. Faculty and third-year residents completed an evaluation card on each student at the end of each shift. Shifts encompassed day, evening, and overnight time periods and included both the adult and pediatric departments.

Our evaluation card contained 10 components used to evaluate the students while in the clinical arena (Figure 1). The card allowed for six grading levels at Fail (0), Marginal (1), Concern (1.5), Pass (2), High Pass (2.5), and Honors (3.0). There was also an option for "Not Evaluated" if the faculty felt they did not have enough information to render an opinion. The reverse side contained space for free-text comments and listing procedures.

During the study period, 25 full-time faculty worked in the ED. All faculty were provided a criterion-based grading scale for the full 10 component evaluation card, based on the six grading levels (Appendix A). We based the criterion-based grading scale on scales used at multiple other institutions. The criterion-based grading scale was reviewed by and with faculty and residents to insure understanding of the scale at faculty meetings and during resident as teacher sessions. Faculty and students were able to access the criterion-based grading scale at any time on the web-based clerkship website. Faculty also received feedback on their grading as compared to all other faculty. The back of the evaluation card provided space for and specifically requested written feedback.

Shift evaluations comprised 65% of the final grade, while a locally developed test provided 25% and adjunct pieces 10% of the final grade. The final grade was determined using criterion-based cutoffs for each of these items. The test (written according to the then current National Board of Medical Examiners question standards) was reviewed for discrimination and reliability by Kuder Richardson (KR)-21, KR-22, and Spearman-Brown statistics. The adjunct pieces included an oral presentation and simulation lab / cadaver lab grades. Demographic data (age, race) were not retained within the database.

Data Analysis

A correlation matrix using Spearman's rho was created using the faculty evaluation components to examine interrelationship between responses (Table 1). We observed a natural clustering between certain evaluation components

Table 1. Correlation matrix between evaluation category average scores, using Spearman's rho. All correlations were significant after Bonferroni correction at p<0.0001.

	Medical knowledge	History taking	Physical exam	Differential diagnosis	Presentation	Patient rapport	Team rapport	Motivation	Procedure skills
Medical knowledge	1.00								
History taking	0.91	1.00							
Physical exam	0.89	0.92	1.00						
Differential diagnosis	0.90	0.89	0.89	1.00					
Presentation	0.83	0.87	0.87	0.88	1.00				
Patient rapport	0.77	0.81	0.81	0.77	0.89	1.00			
Team rapport	0.76	0.79	0..79	0.75	0.86	0.93	1.00		
Motivation	0.77	0.79	0.80	0.76	0.83	0.88	0.89	1.00	
Procedural skills	0.66	0.60	0.66	0.60	0.58	0.59	0.59	0.60	1.00
Documentation	0.73	0.73	0.77	0.75	0.75	0.75	0.74	0.73	0.61

(rho>0.80) in addition to face validity, leading to the establishment of two new variables: clinical skill (composed of a combined average of medical knowledge, physical exam, history, differential diagnosis, and case presentation) and interpersonal skills (combined average of patient rapport, team rapport, and motivation). It should be noted that case presentation also had high correlation with both clinical skills and interpersonal skills; however, it was grouped with the clinical skills variable. Likewise, patient rapport had weaker but still substantial (rho=0.81) correlation with history taking and physical exam; however, the correlations were stronger with the other interpersonal skills, leading us to group patient rapport with team rapport and motivation. Documentation quality and procedural skill did not correlate strongly with other components and were therefore considered in the modeling as separate covariates.

As a sensitivity analysis, to gauge the effect of expected loss of information due to collapsing variables, we constructed separate multiple variable linear regression models using the full model (the weighted average, weighted on number of evaluations, of each of the 10 components) and the reduced model (weighted averages of clinical skill, intrapersonal skill, procedural skill, and documentation quality) in terms of predicting the student's final grade for the rotation.

We tested each model for normality of the residuals via the Shapiro-Wilk test and heteroscedasticity of the residuals with the Breusch-Pagan test. The adjusted R^2 was obtained for each model and 95% confidence intervals (CI_{95}) were calculated. Given that the sample size was constrained due to the number of students completing the rotation during the study period, we did not conduct formal power analysis. However, given the standard "rule of 10" for regression modeling (10 subjects for every degree of freedom included in the linear regression model), and given that the largest model contained 10 covariates, a minimum of 100 students would be required to avoid potentially overfitting the models. Demographic data (age, race) had not been retained within the administrative database and were therefore not available for inclusion in the modeling process. We calculated statistics using Stata 10.1/SE (College Station, TX). An alpha of <0.05 was held to be statistically significant, and we made adjustments for multiple comparisons in the correlation matrix using the Bonferroni method.

RESULTS

Over the two academic years the data were collected, 233 students completed the clerkship. The mean number of evaluations per student was 11.7 (CI_{95} (Poisson exact) 11.3–12.1). Both models satisfied the assumptions of normality and homoscedasticity of the residuals. The full model was significantly predictive of the final grade ($F_{10,222}$=5.96, p<0.0001) and had an adjusted R^2=0.71 (CI_{95}=0.65–0.77). The reduced model, using composite variables reflecting clinical skill and interpersonal skill, was likewise predictive of the final grade ($F_{4,228}$=129.64, p<0.0001) and accounted for a similar proportion of the variance in final grade (adjusted R^2=0.69, CI_{95}=0.63–0.76).

DISCUSSION

Clerkship directors have questioned whether or not evaluators are actually assessing all components of a global rating index of a medical student's performance or whether they are grouping certain aspects of the evaluation together. By looking at the correlation coefficients of our GRS, there is a strong correlation between the clinical skill components of medical knowledge, physical exam, history, differential diagnosis, and case presentation. The interpersonal skills that had a strong correlation were patient rapport, team rapport, and motivation. The final two components that did not show a strong correlation with each other or the other groupings were procedural skills and quality of documentation. These results are similar to those described by Bandiera et al, although the correlation coefficients in this study signify an even stronger correlation.[7]

Although we can debate the benefit of dividing the clinical skills grouping out into the various components to allow for better feedback to the student on areas of strengths and weaknesses, it appears that the attending or resident's evaluation does not vary significantly among these components. It may seem to the educator that the evaluation of a student's physical exam skills is quite different from the student's medical knowledge assessment, but in reality there is a strong correlation. One argument for this correlation is that the well-performing student does well on all components while the poor performing students stumble on all components equally. It is the authors' belief that the evaluator develops a global assessment on the student based on overall clinical and interpersonal skills and then links all the grades together in each of these categories.

A well-suited question may be how to provide faculty with the tools to separate these components. Criterion-based evaluation forms define the different components and even the behaviors/qualities associated with each grade level within the scale for each individual component. Despite the fact that we educated faculty and residents on the components and behaviors/qualities appropriate for a given grade level, our data suggest that the evaluation cards submitted after faculty development had high correlation between these components. Whether or not the faculty or residents give directed feedback on these areas during the end-of-shift evaluation discussion is not known.

A secondary outcome of our study shows that combining these components into a four-component evaluation was just as predictive of the final grade. This finding is consistent with research in other clerkships with longer periods (>4 weeks) of evaluator-student interaction which showed a reduced model being as effective in predicting the final grade as a more detailed evaluation form.[4,6] Therefore, based on these data, clerkship directors should be wary of developing lengthy GRSs for faculty and residents to complete on student's performance. Other means of assessment, such as direct observation, standardized patients, etc. may need to be considered in order to develop a more accurate picture of a student's ability.

LIMITATIONS

This study is limited by the fact that we collected data retrospectively. Within the confines of our medical school grading structure, we were unable to directly assess one evaluation scale against the another. We were unable to assess if verbal feedback differentiated among the different components of the evaluation card. Lastly, the GRS used here is not a validated scale.

CONCLUSION

This study revealed that several components of the evaluation card had a high degree of correlation. This finding calls into question whether a GRS can accurately discriminate between different components. When grouped together into a reduced model containing four components, the evaluation card maintained its predictive level for the final clinical grade. Therefore, when using a GRS for assessment, clerkship directors should evaluate the performance of the GRS in discriminating between components and the feedback provided from the GRS scores.

Address for Correspondence: Kim Askew, MD, Wake Forest School of Medicine, 1 Medical Center Blvd, Winston-Salem, NC 27157. Email: kaskew@wakehealth.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Khandelwal S, Way DP, Wald DA, et al. State of undergraduate education in emergency medicine: a national survey of clerkship directors. *Acad Emerg Med.* 2014;21(1):92-5.
2. Bandiera G and Lendrum D. Daily encounter cards facilitate competency-based feedback while leniency bias persists. *CJEM.* 2008;10(1):44-50.
3. Yarris LM, Fu R, Lamantia J, et al. Effect of an educational intervention on faculty and resident satisfaction with real-time feedback in the emergency department. *Acad Emerg Med.* 2011;18(5):504-12.
4. Pulito AR, Donnelly MB, Plymale M. Factors in faculty evaluation of medical students' performance. *Med Educ*. 2007;41(7):667-75.
5. Metheny WP. Limitations of physician ratings in the assessment of student clinical performance in an obstetrics and gynecology clerkship. *Obstet Gynecol*. 1991;78(1):136-41.
6. Lee M and Wimmers PF. Clinical competence understood through the construct validity of three clerkship assessments. *Med Educ.* 2011;45(8):849-57.
7. Bandiera GW, Morrison LJ, Regehr G. Predictive validity of the global assessment form used in a final-year undergraduate rotation in emergency medicine. *Acad Emerg Med*. 2002;9(9):889–95.

Effect of Doximity Residency Rankings on Residency Applicants' Program Choices

Aimee M. Rolston, MD, MS*
Sarah E. Hartley, MD†
Sorabh Khandelwal, MD‡
Jenny G. Christner, MD§,¶
Debbie F. Cheng*
Rachel M. Caty‖
Sally A. Santen, MD, PhD#

*University of Michigan Medical School, Ann Arbor, Michigan
†University of Michigan Medical School, Department of Internal Medicine, Ann Arbor, Michigan
‡Ohio State University, Department of Emergency Medicine, Columbus, Ohio
§Upstate Medical University, Department of Pediatrics, Syracuse, New York
¶Baylor College of Medicine, Houstion, Texas
‖University of Michigan, School of Public Health, Ann Arbor, Michigan
#University of Michigan Medical School, Department of Emergency Medicine, Ann Arbor, Michigan

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted May 2, 2015; Revision received August 9, 2015; Accepted August 18, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.27343

Introduction: Choosing a residency program is a stressful and important decision. Doximity released residency program rankings by specialty in September 2014. This study sought to investigate the impact of those rankings on residency application choices made by fourth year medical students.

Methods: A 12-item survey was administered in October 2014 to fourth year medical students at three schools. Students indicated their specialty, awareness of and perceived accuracy of the rankings, and the rankings' impact on the programs to which they chose to apply. Descriptive statistics were reported for all students and those applying to Emergency Medicine (EM).

Results: A total of 461 (75.8%) students responded, with 425 applying in one of the 20 Doximity ranked specialties. Of the 425, 247 (58%) were aware of the rankings and 177 looked at them. On a 1-100 scale (100=very accurate), students reported a mean ranking accuracy rating of 56.7 (SD 20.3). Forty-five percent of students who looked at the rankings modified the number of programs to which they applied. The majority added programs. Of the 47 students applying to EM, 18 looked at the rankings and 33% changed their application list with most adding programs.

Conclusion: The Doximity rankings had real effects on students applying to residencies as almost half of students who looked at the rankings modified their program list. Additionally, students found the rankings to be moderately accurate. Graduating students might benefit from emphasis on more objective characterization of programs to assess in light of their own interests and personal/career goals. [West J Emerg Med. 2015;16(6):889–893.]

INTRODUCTION

Choosing a residency program is a stressful and important decision for any medical student. The choice of training program will likely influence their future practice and location.[1] Currently, applicants largely base their decision-making on both 1) personal factors such as geographic location and quality of life; and 2) program factors such as expected clinical experience, curriculum quality, academics, reputation of program, the interview day, and experience with residents and faculty.[2-6] Factors that Emergency Medicine (EM) program directors felt impact applicants' program choices include the interview experience, personal experience with residents, and academic reputation.[5]

Doximity released residency program rankings by specialty in September 2014 with some collaboration from U.S. News and World Report (USNWR). Doximity is a free, HIPAA-compliant online platform for physicians' social networking, collaboration, and education. Using their physician network, Doximity administered a survey asking practicing physicians to "nominate up to 5 residency programs in your medical specialty that offer the best clinical training. Do not consider geography. All nominations will receive the same weight regardless of the order in which you list them."[7,8] More than 17,000 Doximity members responded to the survey, which was conducted between January and July of 2014. Nominations were weighted to account for regional differences in response rates and in the proportion of physicians who are Doximity users.[8] The result was a ranking of the residencies in each of the 20 surveyed specialties. In addition, Doximity also created the Residency Navigator that includes additional information, when available, such as percentage of graduates from a program who specialize, board pass rate, and alumni with peer review articles, grants or clinical trials. However, this objective information is not included in the ranking lists available to the public as the rankings are "based solely on the reputational component." Only Doximity members have access to the majority of the added objective data.[9]

The leaders of the national EM organizations responded to the rankings with concern "about the sampling method chosen for this survey, because we believe it will fail to achieve [the] objective for this survey — to identify America's top EM training programs."[10] Arguably, a survey based on reputation alone cannot objectively measure the quality of hands-on-training and other unique aspects of a residency program (e.g., patient acuity, number of procedures, trauma experience, resident satisfaction). The effects of the Doximity findings, which have both reputational and ranking implications, are not yet known, and they could result in changes to applicants' selections of residency programs. Simply looking at the rank list may bias the candidate when selecting programs for interviews or when ranking programs for the match. Lower ranked programs may suffer the consequences of these rankings by missing quality candidates who may choose not to apply. The objective of this study was to investigate the impact of the Doximity rankings on the program choices made by residency applicants.

METHODS

Survey Design

The survey was developed by educational leaders in undergraduate and graduate medical education and senior medical students, all familiar with the residency application process (content validity). The survey was piloted by 20 residents and faculty and revised for response process validity. This study was determined IRB exempt at all three participating schools.

Survey Content, Administration and Population

The final 12-item survey was sent by email using Qualtrics™ to all fourth year students applying through the National Resident Matching Program at three medical schools in October 2014, just after the release of the Doximity rankings. Student responses were anonymous. Repeated requests were sent by emails weekly to non-responders for three consecutive weeks. The survey initially asked the specialty to which the student applied and whether the student was aware of and looked at the Doximity rankings prior to submitting their application to specific residencies. Students that applied in one of the 20 ranked specialties and who had looked at the rankings were also asked demographic information, how accurate they perceived the rankings on a 100-point scale (0 being not accurate at all and 100 being very accurate), for a narrative to support their score, and whether they added or dropped programs based on the rankings. Additionally, space was provided for students to comment about the rankings.

Data Analysis

Data analysis included descriptive statistics using SPSS (v22, IBM Corp). Comments were analyzed using grounded theory by a single author. The comments were reviewed, codes identified and then grouped based on common themes. Results were summarized based on these themes.

RESULTS

A total of 461 students responded to the survey across all three schools (overall response rate of 75.8%), with 425 students applying in one of the 20 ranked specialties by Doximity (see supplemental Table for distribution of specialties). Forty-seven students applied to EM. Of the 425 students applying in one of the ranked specialties, 58% were aware of the rankings and 72% of those aware looked at the rankings (Figure). The demographics of this sample of applicants who looked at the rankings were: mean age 26 years (range 24–33 years), 50% women, 66% self-identified White, 26% Asian and, 5% Black or African American.

Respondents found the rankings moderately accurate with the mean score for accuracy of 56.7 (SD 20.3, range 0-99) for all students; the accuracy rating of the EM applicants was lower with a mean of 43.3 (SD 23.1, range 9-85) (Table 1a and 1b). Of the 114 students who gave justification for their accuracy score, approximately half of them noted that the Doximity rankings did not include all relevant factors in making a choice on a residency program (N=56, 49%). Student comments included "the culture of a program (learning environment, community, resident community) is not reflected in the Doximity scores" and "[it's] difficult to assess the entire hospital/program based on rankings of different subspecialties. Also difficult to assess patient care vs. research." On the other hand, 22

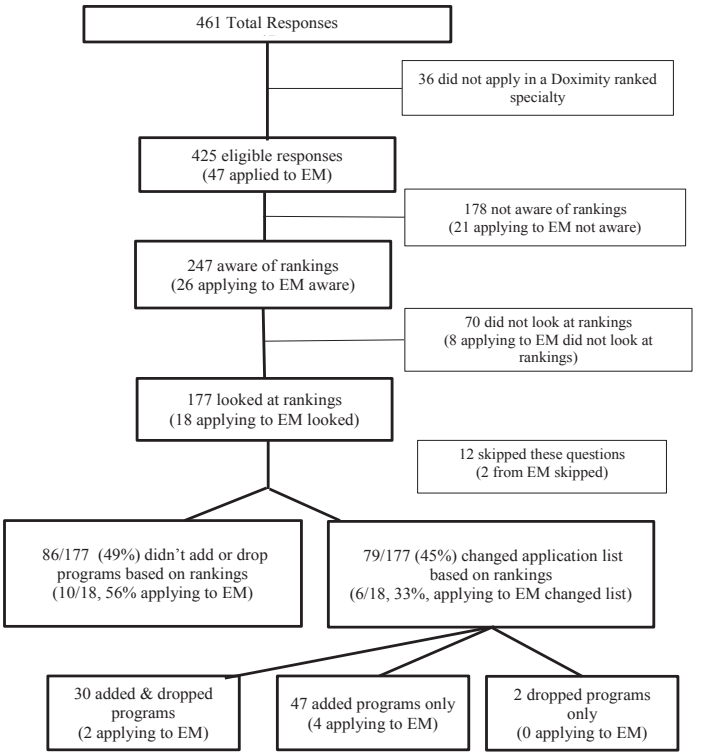


Figure. Flow chart showing number of students aware of rankings, that looked at rankings and that modified the list of programs applied to based on rankings.

Table 1a. Doximity accuracy ratings and modifications to program list.

	Total sample mean (SD; range)	EM only mean (SD; range)
Accuracy rating (n=162; 17)	56.7 (20.3; 0-99)	43.3 (23.1; 9-85)
Mean number of programs added[a] (n=77; 6)	4.32 (3.0; 1-11)	4.83 (3.5; 2-11)
Respondents dropping programs[a] (n=32; 2)	2.88 (2.2; 1-11)	3.00 (1.4; 2-4)

EM, emergency medicine
[a]Scale of "1" – ">10". Note that choice of ">10" programs added or dropped was coded as "11" for purposes of determining the mean.

Table 1b. Factors and reasoning for modifications.

	Total sample number (%)	EM only number (%)
Reasoning for adding or dropping		
Added programs highly ranked	62 (39%)	4 (25%)
Dropped programs lowly ranked	20 (13%)	0 (0%)
Added "safety" programs	25 (16%)	4 (25%)
Added "reach" programs	24 (15%)	2 (13%)
Other	2 (1%)	0 (0%)
Important factors		
Couples matching	24 (15%)	5 (28%)
Geographic location	150 (91%)	17 (94%)
Academic Reputation	146 (89%)	15 (83%)
Personal Connection	99 (60%)	13 (72%)
Choice to stay at home institution	37 (22%)	5 (28%)
Other	18 (11%)	0 (0%)

(19%) comments indicated that the student felt that the ranking correlated with previous conversations, personal experience, student forums, gut feeling and USNWR rankings. Such comments included "[rankings] appear to align with opinions of advisors" and "There aren't any other alternatives, and the top programs seemed to align with common knowledge."

Seventy-nine (45%) of the 177 students who looked at the rankings changed the list of programs to which they applied based on the rankings (Figure). The mean number of programs added and dropped was 4.32 and 2.88, respectively (Table 1a). More specifically, 17% (N=30) both added and dropped programs, 27% (N=47) added programs only, 1% (N= 2) dropped programs only and 49% (N=86) did not add or drop programs, based on the Doximity rankings. Twelve respondents (7%) did not indicate whether they added or dropped programs (Figure). Specifically for students applying to EM, the numbers were similar to the entire sample with a mean number of programs added and dropped of 4.8 and 3.0, respectively.

DISCUSSION

This study found that medical students utilized and reacted to the Doximity residency rankings with a substantial proportion of participants changing their program choices as a result of viewing the rankings. The majority of students who changed their application list added programs only, increasing the number of programs to which they applied and leading to a potential increased cost. In contrast, some students dropped programs, indicating they excluded residencies initially considered. Lastly, some students were aware of the rankings, and either chose not to check the rankings and/or change their application list.

Students found the rankings, on average, to be only moderately accurate. Our analyses did not break down those that added or dropped programs based on their accuracy rating, but, on average, students appear to be reacting to a ranking they view as only moderately accurate. This may be because students use several pieces of information (all to a varying degree) to make their choices and, therefore, are willing to incorporate an only somewhat reliable source as its impact can be modulated against other pieces of information. Alternatively, this may be a result of the fact that no better data exist.

Some of the accuracy concerns provided by students' comments highlight potential methodological issues with the Doximity survey that corroborate with the concerns expressed by the leaders of national EM organizations. These potential methodological issues include construct validity which refers

to whether an indicator measures what it is intended to measure (i.e., the top programs in the country), and measurement validity which refers to the errors that may ensue in the measurement process.[11] First, responses were subjective in nature as many physicians do not have first-hand experience with programs other than their own and the ones they attended. Furthermore, there is also a risk of sampling bias with Doximity's polling methods.[10] The use of a social media website and inclusion of input from only physicians who are members of Doximity excludes the opinions of many and may sway results based on the characteristics of physicians who sign up for a service such as Doximity.[12] The rankings are easily manipulated by programs through encouraging their faculty and alumni to join Doximity and cast votes. Lastly, an additional example calling into question the validity of the instrument is that one of the top ten programs in EM was on probation at the time.[13]

Reputation affects decision-making;[14-16] and Doxmity rankings may be a surrogate for reputation for students. However, it is also important to recognize that a number of factors, beyond reputation, influence medical students' decision-making as they decide which residency programs to which to apply, including objective measurements (such as those included in the Residency Navigator portion of the Doximity report but excluded from the rankings), advising from mentors, and personal reasons. Students also need a means to assess programs and a mechanism to look for specific opportunities that align with their career interests, goals for training, geographical preference, and any influences on family members and personal relationships (such as couples matching).The Doximity rankings could be enhanced by the inclusion of objective data.

Despite substantial research, it is still unclear how we truly make decisions. Emotions and rationality each play a part. For students, the decision to apply to highly ranked programs appeals both to the emotions of success and competition as well as to rationality, which encourages them to choose pathways more likely to lead to success. Other influences on decision-making are biases and heuristics, which are unconscious routines to cope with the complexity inherent in most decision-making.[17] One heuristic is the assumption that a higher ranked program will provide better training. Additionally, anchoring may lead to weighing certain pieces of information too heavily in the decision-making process. Similarly, confirmation bias leads people to ignore evidence that contradicts their preconceived notions. The rankings can play into these biases and, as a result, students may allow decisions to be based on rankings as a surrogate for quality of training.

Perhaps the best way to aid applicants is to move away from rankings and, instead, provide and focus on objective data about programs that students can judge in light of their own interests, career goals and personal preferences. The concept of providing students a resource such as the Residency Navigator to pull data together might be useful without an overall "ranking." A process to help programs demonstrate data relevant to finding the right "fit" for a residency and other objective data might include setting (rural vs. urban, public vs. private), academic, research or community focused, board certification scores, in-service training examinations, selectivity, percent of residents progressing on track for specialty milestones, numbers of procedures performed, measurements on the annual ACGME program evaluation, and accreditation and hospital metrics such as Quality Leadership Awards. This information could provide a set of metrics to characterize programs in a transparent fashion. Certain resources serve as a precedent for this.[18] The Residency Navigator component of the Doximity study attempted to begin such characterization, but it was unfortunately overshadowed by the fanfare of the published rankings. Unless residency programs agree to publish objective data to be used by applicants for best fit, published rankings, such as the one by Doximity, may gain more acceptance and importance over time despite their shortcomings.

LIMITATIONS

The sample of students applying to EM was small, but they appear similar to the general population in using the Doximity rankings to determine their application list. This small sample size of EM applicants may limit generalizability and future studies should expand the EM sample to other schools. Additionally, surveyed students did not represent all geographic areas, further limiting generalizability. Lastly, recall bias is a potential limitation to these results, as students may not remember exactly how the rankings affected their list. We attempted to limit this effect by surveying students only a few weeks after the initial opportunity for application submission (i.e., September 15).

CONCLUSION

The Doximity residency rankings by specialty influenced the programs to which fourth year medical students chose to apply. On average, students viewed the rankings as only moderately accurate. These rankings were based on reputation data and did not include objective measures. Rankings are often perceived as offering an objective reality of what is "best." However, what is best for one applicant may be quite different than what is best for another. Residency applicants would likely be better served by providing students with and focusing them on objective program data that they can consider in terms of their own career and personal goals.

Address for Correspondence: Aimee Rolston, MS, University of Michigan Medical School, 5121 Medical Science Building I,1137 Catherine St. Ann Arbor, MI 48109-5808 Email:aimeerol@umich.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. American Association of Medical Colleges. Physician Specialty Data Book. 2012; Available at: https://http://www.aamc.org/download/313228/data/2012physicianspecialtydatabook.pdf. Accessed Oct 25, 2014.
2. Davydow D, Bienvenu OJ, Lipsey J, et al. Factors influencing the choice of a psychiatric residency program: A survey of applicants to the Johns Hopkins residency program in psychiatry. *Acad Psychiatry.* 2008;32(2):143-6.
3. Lee J, Alfieri M, Patel T, et al. Choosing family medicine residency programs: What factors influence residents' decisions? *Can Fam Physicians.* 2011;57(3):e113-e121.
4. DeSantis M and Marco CA. Emergency medicine residency selection: Factors influencing candidate decisions. *Acad Emerg Med.* 2005;12(6):559-61.
5. Love JN, Howell JM, Hegarty CB, et al. Factors that influence medical student selection of an emergency medicine residency program: Implications for training programs. *Acad Emerg Med.* 2012;19(4):455-60.
6. Stefanidis D, Miles WS, Greene FL. Factors influencing residency choice of general surgery applicants—How important is the availability of a skills curriculum? *J Surg Educ.* 2009;66(6):325-29.
7. Doximity. About Doximity. Available at: https://http://www.doximity.com/about/company. Accessed Oct 25, 2014.
8. Harder B and Heilbrunn E. Doximity's 'Residency Navigator' injects transparency into GME. 2014; Available at: http://health.usnews.com/health-news/blogs/second-opinion/2014/09/10/doximitys-residency-navigator-injects-transparency-into-gme. Accessed Oct 25, 2014.
9. Doximity. Doximity residency navigator: Our residency research methodology. Available at: https://s3.amazonaws.com/s3.doximity.com/mediakit/Doximity_Residency_Navigator_Survey_Methodology.pdf. Accessed Nov 2, 2014.
10. Joint EM Organization Letters Regarding Doximity Medical Student Survey. 2014; Available at: http://www.cordem.org/i4a/pages/index.cfm?pageid=3359. Accessed Nov 1, 2014.
11. Ioannidis JPA, Patsopoulos NA, Kavvoura FK, et al. International ranking systems for universities and institutions: a critical appraisal. *BMC Medicine.* 2007;5:30.
12. Rosenau AM, Mercer M, Mitchell M, et al. Letter to Mr. Ben Harder, USNWR. 2014; Available at: http://saem.org/docs/default-source/default-document-library/finalusnewsletter.pdf?sfvrsn=2. Accessed Oct 25, 2014.
13. Accreditation Council for Graduate Medical Education. Public Programs or Institutions with Probationary Status. 2014; Available at: https://http://www.acgme.org/ads/Public/Reports/Report/9. Accessed Oct 25, 2014.
14. Meshi D, Biele G, Korn CW, et al. How expert advice influences decision making. *PLoS One.* 2012;7(11):e49748.
15. Harvey N and Fischer I. Taking advice: Accepting help, improving judgment, and sharing responsibility. *Org Behav Human Decision Proc.* 1997;70(2):117-33.
16. Cole JR and Lipton JA. Reputations of American medical schools. *Social Forces.* 1997;55(3):662-84.
17. Kahneman D, Lovallo D, Sibony O. Before you make that big decision...: dangerous biases can creep into every strategic choice. Here's how to find them - before they lead you astray. *Harv Bus Rev.* 2011;89:50-60.
18. Association of Professors of Gynecology and Obstetrics. Residency directory. Available at: https://http://www.apgo.org/student/residency.html. Accessed Oct 25, 2014.

Introducing Medical Students into the Emergency Department: The Impact upon Patient Satisfaction

Christopher Kiefer, MD*
Joseph S. Turner, MD†
Shelley M. Layman, MPA*
Stephen M. Davis, MPA, MSW*
Bart R. Besinger, MD†
Aloysius Humbert, MD†

*West Virginia University School of Medicine, Department of Emergency Medicine, Morgantown, West Virginia
†Indiana University School of Medicine, Department of Emergency Medicine, Indianapolis, Indiana

Section Editor: Douglas S. Ander, MD
Submission history: Submitted May 1, 2015; Revision received September 4, 2015; Accepted September 26, 2015
Electronically published November 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27255

Introduction: Performance on patient satisfaction surveys is becoming increasingly important for practicing emergency physicians and the introduction of learners into a new clinical environment may impact such scores. This study aimed to quantify the impact of introducing fourth-year medical students on patient satisfaction in two university-affiliated community emergency departments (EDs).

Methods: Two community-based EDs in the Indiana University Health (IUH) system began hosting medical students in March 2011 and October 2013, respectively. We analyzed responses from patient satisfaction surveys at each site for seven months before and after the introduction of students. Two components of the survey, "Would you recommend this ED to your friends and family?" and "How would you rate this facility overall?" were selected for analysis, as they represent the primary questions reviewed by the Center for Medicare Services (CMS) as part of value-based purchasing. We evaluated the percentage of positive responses for adult, pediatric, and all patients combined.

Results: Analysis did not reveal a statistically significant difference in the percentage of positive response for the "would you recommend" question at both clinical sites with regards to the adult and pediatric subgroups, as well as the all-patient group. At one of the sites, there was significant improvement in the percentage of positive response to the "overall rating" question following the introduction of medical students when all patients were analyzed (60.3% to 68.2%, p=0.038). However, there was no statistically significant difference in the "overall rating" when the pediatric or adult subgroups were analyzed at this site and no significant difference was observed in any group at the second site.

Conclusion: The introduction of medical students in two community-based EDs is not associated with a statistically significant difference in overall patient satisfaction, but was associated with a significant positive effect on the overall rating of the ED at one of the two clinical sites studied. Further study is needed to evaluate the effect of medical student learners upon patient satisfaction in settings outside of a single health system. [West J Emerg Med. 2015;16(6):894–898.]

INTRODUCTION

In 2006, in response to the growing shortage of physicians in the United States, the Association of American Medical Colleges advocated for a 30% increase in enrollment at Liaison Committee on Medical Education (LCME) accredited institutions by the year 2015.[1] As of 2012 LCME schools were

on target to reach the goal.[2]

In addition to increased enrollments, demand for emergency medicine (EM) training venues is also created by the increasing number of required EM clerkships at LCME schools, having risen from 33% to 52% of schools between 2003 and 2010.[3,4] Since most medical students currently complete their EM clerkship in large, academic, tertiary care hospitals where resident physicians are also present, institutions may need to increasingly utilize alternative clinical settings, including community-based emergency departments (EDs) as training sites for medical students.

As community EDs begin to host an increasing number of medical student rotators, concerns have arisen regarding the impact upon patient experience and satisfaction. Prior work has illustrated that although community-based EDs may have lower patient volumes than the primary university site, they are viable teaching sites, as students rotating at community-based sites have a significantly higher number of patients evaluated per shift, a significantly higher number of procedures per shift, and gave higher clinical teaching scores to attending physicians.[5]

Attending physicians practicing in community-based EDs are facing increased pressure to perform well on patient satisfaction surveys. The introduction of value-based purchasing (VBP) by the Center for Medicare Services (CMS) has tied portions of hospital reimbursement to patient satisfaction. The results of patient satisfaction surveys may be used by hospitals to make credentialing decisions regarding individual physicians, make bonus payments to physicians, as well as to determine the contract status for an entire group of physicians.[6] It has been hypothesized that attending physicians in this setting may be hesitant to participate in teaching experiences, fearing the negative impact that the presence of learners will have upon patient satisfaction.[7]

While it appears that community EDs are viable venues for clinical education, physicians in this setting are also interested in maintaining high patient satisfaction. To our knowledge, there has been no prior work that evaluates the relationship between patient satisfaction and the introduction of medical students to the community ED setting. This study seeks to examine the effect of introducing medical students into two community-based EDs upon performance on institutional patient satisfaction surveys.

METHODS

At Indiana University, EM is a required clerkship for all fourth-year medical students that may be completed at a variety of clinical sites – ranging from large, tertiary academic centers to smaller community-based ED settings throughout Indiana. Two community-based EDs in the Indiana University Health (IUH) system, IUH North (Site A) and IUH Saxony (Site B), were chosen for analysis in this study given that prior to the start date of medical students in the ED, there were no medical students or residents present in the hospital. During the study period, both sites were staffed using a single coverage model, where there was a single American Board of Emergency Medicine-certified physician working at all times. While there was not a medical student present at either site during all shifts, all physicians at both sites worked with medical students when a student was present. The study was designed as a retrospective cohort of existing patient satisfaction data collected by both sites during the study period as part of standard ED operations.

Patient satisfaction surveys were administered to patients discharged from the EDs in the study by a third party, National Research Corporation (NRC) Picker. The surveys were administered by NRC Picker to discharged patients from both ED sites via US mail as per the pre-existing contract between the institution and NRC Picker. No additional patient satisfaction surveys were administered during the study period. Surveys were returned to NRC Picker by individual patients at their discretion. If an individual patient did not answer a question on the survey, we excluded that data point from analysis for that particular question.

Survey results were reported to the clinical sites on a monthly basis during the study period as a percentage of positive response to each question. The standard report provided to both clinical sites reported the percentages of positive response for adult, pediatric, and all patients.

We analyzed surveys for the seven months before and after the introduction of medical students at each clinical site with the study powered to detect a 15% difference in positive response rate. Medical students were introduced at Site A in March 2011 and at Site B in October 2013. As such, studies collected from August 2010-February 2011 and March 2011-October 2011 formed the pre- and post-medical student cohorts at Site A with surveys from February 2013-August 2013 and October 2013-April 2014 forming the pre- and post-medical student cohorts at Site B.

For purposes of the study, two questions on the survey were subjected to data analysis–"Would you recommend this ED to your friends and family?" and "How would you rate this facility overall?"–where a positive response was considered a 9 or 10 on a scale of 1-10 or a score of excellent prior to January 2011 (Site A only). These questions were chosen for analysis as they represent the primary outcome measures reported by the institution as part of CMS VBP and the results to these questions are publicly reported. "Definitely yes" was considered a positive response for the "would you recommend" question. Responses of 9, 10, or excellent (prior to January 2011) considered as a positive response for the "how would you rate" question. The change from a categorical response "excellent/definitely yes" to a continuous response "9 or 10" represented a change in the survey response options implemented by the survey vendor. As the survey results before and after this alteration were reported as a percentage of positive response, the change in terminology does not

represent a change in the primary outcome, and as such, we included both survey sets in the analysis.

We compared percentages of positive response for the periods before and after the introduction of medical students using a chi-square analysis. This study was deemed exempt by the institutional review board at the Indiana University School of Medicine.

RESULTS

For Site A, 224 surveys were returned in the seven months prior to the introduction of medical students and 520 surveys returned in the seven months following the introduction of medical students. For Site B, there were 247 surveys returned prior to the introduction of the students with 224 surveys returned following the introduction of medical students. The survey response rate was 22.8% for Site A for January 2011-October 2011and 17.2% for Site B for the entirety of the study period. Response rate data prior to January 1, 2011 for Site A was not available. The surveys in this study were administered by a third party; only the total number of surveys returned and the rate of return were reported to the institutions. As such, we were unable to directly ascertain the reasons behind the increased total number of surveys at Site A in the post-medical student period.

Two individuals at Site A and three individuals at Site B did not answer the "would you recommend" question in the pre-student period. Following the introduction of students, two individuals at Site A and six individuals at Site B did not complete the "overall rating" component of the survey.

For Site A, we were unable to detect a statistically significant difference in the percentage of positive response for the "would you recommend" question when the adult (p=0.549), pediatric (p=0.284), or all-patient (p=0.238) groups were analyzed (Table 1). For the same query at Site B, we were again unable to detect a statistically significant difference in the percentages of patients giving a positive response to the "would you recommend" question in the adult (p=0.353), pediatric (p=0.758), or all patients (p=0.756) (Table 2).

Regarding the "overall rating" component of the survey, when we analyzed all patients at Site A, the percentage of patients giving an overall rating of "excellent" or "9/10 and 10/10" increased from 60.3 to 68.2 (p=0.038). However, when broken down into subgroups, we were unable to detect a statistically significant difference between the adult (p=0.347) or pediatric (p=0.062) groups (Table 1). For the same measure at Site B, we were unable to detect a statistically significant difference in the adult (p=0.738), pediatric (p=0.554), or all patient (p=0.476) groups (Table 2).

Table 3 illustrates the pooled patient satisfaction data from both sites. For the "would recommend" question, we were unable to detect a statistically significant difference in the percentage of positive response in the adult (p=0.976), pediatric (p=0.203), or all patient groups (p=0.333). Finally, when analyzing the pooled data for the "overall rating" question, we were unable to detect a statistically significant difference in the before or after medical student groups in the adult (p=0.817), pediatric (p=0.791), or all patient (p=0.625) groups.

DISCUSSION

Patient satisfaction is becoming increasingly important for the practicing emergency physician. To date, no prior study has analyzed the effect of medical student learners on patient satisfaction in the ED setting. In both EDs, there were no medical students or residents present anywhere in the hospital prior to the start date of medical students. While most teaching settings do not have a discrete start date for learners in their facility, the lack of students prior to a certain time point allows for a direct evaluation of the impact of medical students upon patient satisfaction scores.

The addition of learners to a clinical environment adds an additional step in the patient's visit to the ED where the student first evaluates the patient and subsequently presents the findings to an attending physician. Previous work has illustrated that while medical students reduce the time to medical provider, the total length of stay for patients seen by a medical student in the ED is increased by an average of 24 minutes.[8] Total length of stay is certainly an important metric that many EDs follow closely; however, reduced door-to-provider time has been associated with increased patient satisfaction scores.[9] Our study did not measure door-to-provider time directly; however, in a single coverage ED, it is possible that medical students impact this positively and may be an area of additional study.

In the past, patients have generally viewed the involvement of medical students in their care positively. Prislin et al. evaluated patient perceptions of medical student participation in a family medicine clerkship at both community-based and tertiary academic clinics–89% of surveyed patients responded that being seen by a medical student was an enjoyable experience and 77% of patients responded that they felt medical student participation improved the quality of care they received.[10] More recently, a Colombian group found that the introduction of medical students into the inpatient setting improved patient perception of quality of care and overall satisfaction.[11] While our study failed to detect a significant positive difference in most of our measured outcomes, the lack of a negative effect may be reassuring for community-based physicians considering becoming involved in medical student education. Given the increasing demand for medical student education sites and increasing viability of community-based EDs as clinical teaching sites, our results suggest that the presence of medical students does not affect patient satisfaction scores, and patient satisfaction alone should not be considered a barrier to introducing medical students into clinical venues.

LIMITATIONS

Our study has several limitations. First, this study was

Table 1. Percentage of positive responses to the "would you recommend" and "Overall rating" items at Site A before and after the introduction of medical students.

	"Would you recommend?"		"Overall rating"	
	Before students	After students	Before students	After students
Pediatrics	84.1 (n=113)	79.3 (n=256)	50.9 (n=114)	61.3 (n=253)
Adult	84.4 (n=109)	81.9 (n=264)	70.0 (n=110)	74.7 (n=265)
Overall	84.2 (n=222)	80.6 (n=520)	60.3 (n=224)	68.2 (n=518)*

*Differences in percentages were found to be statistically significant ($p<0.05$).

Table 2. Percentage of positive responses to the "would you recommend" and "Overall rating" items at Site B before and after the introduction of medical students.

	"Would you recommend?"		"Overall rating"	
	Before students	After students	Before students	After students
Pediatrics	84.85 (n=99)	83.1 (n=71)	81.82 (n=99)	85.29 (n=68)
Adult	83.11 (n=148)	86.93 (n=153)	83.45 (n=145)	84.87 (n=152)
Overall	83.98 (n=247)	85.02 (n=224)	82.64 (n=244)	85.08 (n=220)

Table 3. Combined percentages of positive responses for both clinical sites before and after the introduction of medical students.

	"Would you recommend?"		"Overall rating"	
	Before students	After students	Before students	After students
Pediatrics	84.85 (n=212)	80.13 (n=327)	65.27 (n=213)	66.38 (n=321)
Adult	83.66 (n=257)	83.75 (n=417)	77.65 (n=255)	78.41 (n=417)
Overall	84.08 (n=469)	81.93 (n=744)	71.95 (n=468)	73.23 (n=738)

performed at two sites within the same health system and used only NRC Picker survey results, and the results may not be generalizable. The analyzed outcomes assessed patient satisfaction with their entire visit. We believe that this is an appropriate measure of patient satisfaction as they represent the primary outcome measures reported by the institution as part of CMS VBP and therefore are meaningful outcomes to the institution and practicing physicians. However, multiple possible confounders could affect the patient's impression of the entire visit. We also did not evaluate individual physician satisfaction scores, though this is an area for potential, future study. This study analyzed a relatively short time period, and does not evaluate larger trends in satisfaction scores. We acknowledge that many health systems are applying service initiatives to increase scores, and this "snapshot" does not evaluate the effect of such efforts.

Due to the fact that patient satisfaction surveys are de-identified, we were unable to focus our analysis on only those patients evaluated by a medical student. Although creation of a separate survey instrument with similar queries would allow for analysis of patients evaluated by a medical student, our study is, to our knowledge, the first to analyze the impact of introducing learners into a clinical environment upon data used for CMS reporting and quality metrics. Finally, our study was powered to detect a 15% difference. While this is a sizable difference, it was chosen in part because Site B is a newly opened hospital, with a limited time of pre-learner data. This limits our ability to detect small differences over a long period of observation.

CONCLUSION

This study suggests that introducing medical students into a community ED does not have a significant impact on patient satisfaction scores. With increasing emphasis on patient satisfaction, the results of this study suggest that sites considering participating in medical student training should be assured that students do not have a negative impact upon patient satisfaction.

ACKNOWLEDGMENT

The authors wish to acknowledge Ms. Joy Graves-Rust, Coordinator of Patient Satisfaction Measures at IUH North Hospital for her hard work and assistance in obtaining the patient satisfaction data used in the study.

Address for Correspondence: Christopher Kiefer, MD, West Virginia University School of Medicine, PO Box 9149, 7413B HSS, 1 Medical Center Dr., Morgantown, WV 26506. Email: ckiefer@hsc.wvu.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Association of American Medical Colleges. AAMC Statement on the Physician Workforce. 2006. Available at: https://www.aamc.org/download/55458/data. Accessed Apr 28, 2015.
2. Association of American Medical Colleges. AAMC Physician Workforce Policy Recommendations. 2012. Available at: https://www.aamc.org/download/304026/data/2012aamcworkforcepolicyrecommendations.pdf. Accessed Apr 28, 2015.
3. Khandelwal S, Way DP, Wald DA, et al. State of undergraduate education in emergency medicine: a national survey of clerkship directors. *Acad Emerg Med*. 2014;21(1):92-5.
4. Wald DA, Manthey DE, Kruus L, et al. The state of the clerkship: a survey of emergency medicine clerkship directors. *Acad Emerg Med*. 2007;14(7):629-34.
5. deLahunta EA and Bazarian J. University and community hospital medical student emergency medicine clerkship experiences. *Acad Emerg Med*. 1998;5(4):343-6.
6. American College of Emergency Physicians (ACEP). Patient satisfaction surveys. Policy Statement. *Ann Emerg Med*. 2011;57(3):313.
7. Coppola LM, Reed KL, Herbert WN. Comparison of patient attitudes and provider perceptions regarding medical student involvement in obstetric/gynecologic care. *Teach Learn Med*. 2014;26(3):239-43.
8. DeLaney M, Zimmerman KD, Strout TD, et al. The effect of medical students and residents on measures of efficiency and timeliness in an academic medical center emergency department. *Acad Med*. 2013;88(11):1723-31.
9. Handel DA, French LK, Nichol J, et al. Associations between patient and emergency epartment operational characteristics and patient satisfaction scores in an adult population. *Ann Emerg Med*. 2014;64(6):604-8.
10. Prislin MD, Morrison E, Giglio M, et al. Patients' perceptions of medical students in a longitudinal family medicine clerkship. *Fam Med*. 2011;33(3):187-91.
11. Esguerra R, Toro J, Ospina JM, et al. The transition to a teaching hospital: patient satisfaction before and after the introduction of medical students. *Med Teach*. 2014;36(8):710-4.

Teaching Emotional Intelligence: A Control Group Study of a Brief Educational Intervention for Emergency Medicine Residents

Diane L. Gorgas, MD
Sarah Greenberger, MD
David P. Bahner, MD
David P. Way, MEd

Ohio State University College of Medicine, Department of Emergency Medicine, Columbus, Ohio

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted May 1, 2015; Revision received August 3, 2015; Accepted August 18, 2015
Electronically published November 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.27304

Introduction: Emotional Intelligence (EI) is defined as an ability to perceive another's emotional state combined with an ability to modify one's own. Physicians with this ability are at a distinct advantage, both in fostering teams and in making sound decisions. Studies have shown that higher physician EI's are associated with lower incidence of burn-out, longer careers, more positive patient-physician interactions, increased empathy, and improved communication skills. We explored the potential for EI to be learned as a skill (as opposed to being an innate ability) through a brief educational intervention with emergency medicine (EM) residents.

Methods: This study was conducted at a large urban EM residency program. Residents were randomized to either EI intervention or control groups. The intervention was a two-hour session focused on improving the skill of social perspective taking (SPT), a skill related to social awareness. Due to time limitations, we used a 10-item sample of the Hay 360 Emotional Competence Inventory to measure EI at three time points for the training group: before (pre) and after (post) training, and at six-months post training (follow up); and at two time points for the control group: pre- and follow up. The preliminary analysis was a four-way analysis of variance with one repeated measure: Group x Gender x Program Year over Time. We also completed post-hoc tests.

Results: Thirty-three EM residents participated in the study (33 of 36, 92%), 19 in the EI intervention group and 14 in the control group. We found a significant interaction effect between Group and Time ($p \leq 0.05$). Post-hoc tests revealed a significant increase in EI scores from Time 1 to 3 for the EI intervention group (62.6% to 74.2%), but no statistical change was observed for the controls (66.8% to 66.1%, $p=0.77$). We observed no main effects involving gender or level of training.

Conclusion: Our brief EI training showed a delayed but statistically significant positive impact on EM residents six months after the intervention involving SPT. One possible explanation for this finding is that residents required time to process and apply the EI skills training in order for us to detect measurable change. More rigorous measurement will be needed in future studies to aid in the interpretation of our findings. [West J Emerg Med. 2015;16(6):899–906.]

INTRODUCTION

Emotional intelligence (EI) is a complex construct involving the perception, processing, regulation and management of emotions.[1] Mayer and Salovey define EI "as a four dimensional construct, comprising the abilities to perceive emotion, use emotion to facilitate thought,

understand emotions, and manage emotion."[2] EI is a construct distinctly different from personality or general intelligence, and has been shown to be a practical predictor of job performance within the most competitive professions such as business, engineering, or medicine.[3-5]

EI is thought to involve four distinct measurable skill sets: self-awareness, self-management, social awareness, and relationship management.[9] While some debate exists as to whether EI is a state (learned) or trait (innate), we conceive of EI as a fluid construct, susceptible to improvement or decline.[7-10]

Although EI is a distinctly different construct than cognitive intelligence (sometimes expressed as intelligence quotient (IQ), and derived from a score on a mental abilities test), the two are thought to have a complex interaction when involved with decision-making in an emotionally charged environment.[1-3] In an emotionally charged environment, individuals with both high EI and IQ are thought to be able to mediate the effects of their own, and others' emotions in order to make sound judgments.[1]

EI and Physician Development

For physicians, EI's pattern of change over time runs counter to change for individuals in other fields.[11-14] Some studies have demonstrated that physician EI deteriorates over time and as training progresses, while professionals in other fields tend towards improvements in EI over time and with experience.[12,14,15] Deterioration in physician EI has been attributed to de-sensitization effects from medical training, de-personalization of the patient-physician relationship, and burnout.[1,11,12,15] Burnout, which can lead to premature career termination, is a common problem in medicine with reported rates as high as 50% in medical students, 75% for residents, and 60% in practicing physicians.[11,15]

A growing body of literature examines the relationship of EI to resident physician wellness, how it relates to resilience during training, and how it impacts career success including career satisfaction and career longevity.[11,12,14,15] EI in and of itself has been shown to be a predictor of success in residency training.[11] Residents with low EI are ill-equipped to cope with the stressors of training and practice, and tend to leave the profession earlier.[5,11,15,16] Residents with high EI have been shown to have more pro-social behavior, better academic performance, better relationships with patients, and longer careers.[11,15]

Strategies for Teaching EI

Teaching strategies to improve an individual's EI vary from simple training in identifying nonverbal cues for emotions, to more advanced exercises in self-reflection and identification of situational stressors.[9,10,17,18] Workplace interventions include education about burnout, workload modifications, increasing the diversity of work duties, stress management training, mentoring, team-building, communication skills, and wellness workshops.[19-21] The ideal intervention for professionals would effectively improve EI in the workplace without huge investments of time and resources.

Successful EI training programs have been conducted at both the undergraduate and graduate medical education levels, suggesting that interventions can be effective during these periods of physician development.[17,20-23]

Simulated patient encounters with feedback have been used to teach individuals, whereas small-group discussion formats have been used to develop EI collectively across groups.[20,22-24] Some studies have shown that female students are more amenable to and gain more from EI training than their male counterparts.[23] Some EI training has yielded mixed or ineffective results, including some with considerable expenditures of effort and time.[16] Generally, the most effective training has been coupled with clinical experience.[17,22]

Measuring EI

There are numerous instruments and scales that purport to measure EI or components of EI. MacCann et. al. reviewed four primary EI instruments including the Emotional Quotient Inventory (EQ-i), the Schutte Self-Report Inventory (SSRI), the Hay 360 Emotional Competence Inventory (ECI), and the Mayer, Salovey, Caruso EI Test (MSCEIT).[25] The instruments are based on slightly varied conceptualizations of the EI construct. The MSCEIT treats EI as a trait or ability, while the other three are considered mixed models because some of the items resemble those found on conventional personality tests, while others are more related to skills or competencies.[25] The instruments vary widely on the composition of their parts or subscales, response options, tasks, stimuli, and methods for scoring.[25] All four EI assessment instruments are comprehensive, complex, and time-consuming.[25]

Purpose

The purpose of this study was to measure the effects of a brief educational intervention designed to improve the general EI of EM residents. Specifically, we taught a group of residents one EI-related skill called social perspective-taking to see if it would improve their EI as operationalized by scores on a 10-item sample from the Hay 360 ECI.[26,27]

METHODS

This study was conducted at a large, academic, urban EM residency training program in the Midwest, during academic year 2011-12. The residency is a three-year program, with 12 residents per class, post-graduate year (PGY) 1-3. Residents at all three levels participated in the study, which was reviewed and approved by our institution's behavioral sciences institutional review board.

Study Design

We used a pre-test, post-test, six-month follow-up randomized control group design. EM residents were assigned to either an intervention (experimental) group or a control group. The assignments were random, but stratified to assure

that both groups had equal numbers of residents from each class and each gender. Both groups were given the sample Hay 360 ECI as a pre-test. The intervention group completed the EI training session and completed the sample Hay 360 ECI as a post-test. Both the intervention and control groups completed the sample Hay 360 ECI as a six-month follow-up assessment.

Educational Intervention

As are most EM residency curricula, ours is densely packed with compulsory content material. Subsequently, our EI intervention and assessments needed to fit into two conference sessions of two hours in length, offered six months apart and alternated with another program activity. Accordingly, our intervention group participated in the EI training in the first two-hour session, and the control group received the training six months later in the second two-hour session.

Time restrictions required us to limit instruction to one sub-skill from one of the primary EI skills, social awareness. We focused on social awareness because of its importance in the clinical setting, and because of its relationship to communication. Social awareness encompasses the sub-skills of empathy, organizational awareness, and social perspective taking (SPT). SPT is a skill related to understanding another individual's viewpoint. Someone with skill in SPT can "accurately infer the thoughts and feelings of others."[4,5,18,28,29] SPT has been shown to predict better communication skills in practitioners and trainees in both nonmedical and medical careers.[5,28-30] We hypothesized that the associated sub-skill of SPT could be taught within our available time limits and that focus on this particular skill would lead to measureable results.

We would have preferred to use the Hay 360 ECI as our measure of EI for three primary reasons. First, the conceptual framework of the instrument was most consistent with our conceptual framework of EI. Second, the assessment uses scenarios as prompts and partial credit scoring, which fit with our repeated measures research design. In other words, we felt that this format would be the least likely to yield a practice effect. Third, the ECI included subscales most directly related to our instruction: empathy, social awareness and SPT.

However, due to cost in time and money, we were restricted to using a measure that was brief, and freely available. Fortunately, we were able to identify a short, sample version of the Hay 360 ECI called The Hay 360 EI Quiz. This was available on the publisher's website, and it took less than 10 minutes for the residents to complete (see Appendix A).[31]

At the beginning of the first EI intervention session, both groups completed the 10-item sample Hay 360 ECI assessment. The control group was then dismissed to a separate conference room for an unrelated residency program activity, while the treatment group remained for the EI intervention.

The EI intervention was modeled after similar programs for physicians and trainees in internal medicine and family medicine.[15,20] Due to time limitations, we chose to focus our training on two very specific components of EI, compassion and SPT.[4,28,32] The session was introduced with a lecture covering basic EI vocabulary and concepts, a description of environmental stressors common to EM residents, and then a brief description about the intended benefits of EI training activities. A video of an interview with Daniel Goleman a leading authority on the topic of EI, and a video of a TED-Talk lecture delivered by Daniel Goleman on the topic of compassion were interspersed into the introductory lecture.[5,32] The lecture was followed by a series of four case scenarios, each involving a person in distress. The cases included a list of suggested actions that an external observer might take, including the reasons why they might take that action.

The first two cases were presented and discussed with the entire intervention group. Participants were asked questions about the cases to guide them to identify the source of distress, the cause of the distress, potential environmental factors, and how the perspective one takes might impact their response to the situation. Two subsequent cases were discussed in facilitated small groups. At the end of the session, each small group presented their case analysis, their response, and a defense of their response. The session concluded with a debriefing in which the "best" responses to each case were identified along with explanations. (See Appendix B for more thorough description of the program.)

The 10-item sample Hay 360 ECI survey was administered to the intervention group at the conclusion of the intervention session.[31] Over the following six months, during their regular clinical work in the emergency department, residents in both groups received regular clinical performance feedback including that which involved ACGME Milestones in interpersonal communication. Both the intervention group and control group retook the 10-item sample Hay 360 ECI survey at the end of the six-month period.

Data Analysis

The overall study design involved four independent variables: group (intervention vs. control); gender (female vs. male); program year (1 vs. 2 and 3); and time (1: pre-intervention, 2: post-intervention, and 3: six-month follow up). There was one dependent variable, the sample Hay 360 ECI as a measure of EI. The preliminary omnibus analysis was a 4-way analysis of variance (ANOVA) with one repeated factor (time). Since both groups (intervention and controls) were not measured at all three time periods, we dropped time 2 from this analysis.

Due to the repeated measure nature of our design, we used partial eta^2 for computations of effect size, as suggested by Brown.[33] We further analyzed the significant two-way interaction of group x time with an analysis of the simple effects of groups separately using paired t-tests.[34] Since the simple effect was significant for the intervention group only, we conducted a 1-way repeated measures ANOVA as a post-hoc analysis of their sample Hay 360 ECI score change over time.

RESULTS

Thirty-three EM residents participated in the study (33 of 36, 92%), 19 in the EI intervention group and 14 in the control group. Participants were evenly split by gender (16 women, 17 men), and by program year (13 PGY-1, 9 PGY-2, and 11 PGY-3) (Table 1). Due to scheduling conflicts, we ended up with slightly more PGY-3 residents in the intervention group. So to verify the equivalence of the groups at baseline, we ran an independent t-test and found the two groups to be statistically equivalent, (Intervention Mean=62.6; Control Mean=66.8; t= -1.02; df=31; p=0.316).

The four-way ANOVA with one repeated measure resulted in a significant interaction effect between group and time ($F=7.16_{(1,21)}$; $p \leq 0.05$). No other main effects or interactions were statistically significant (Tables 2, 3a, and 3b). The means plot (Figure 1) of this interaction shows a marked increase in EI score for the intervention group and no detectable change in EI score for the control group. The partial eta^2 effect size of the group x time interaction was 0.25, which is considered large.[35]

We analyzed the significant interaction with an analysis of simple effects. This was accomplished through paired t-tests of each group's pre- and post-EI score for each group separately. We found that the change in EI mean scores for the intervention group from Time 1 to Time 3 (62.6 to 74.2) was significant ($t_{1,18}= -3.54$; $p \leq 0.01$) while the change over the same time period for the control group (66.8 to 66.1) was not ($t_{1,13}=0.31$; $p=0.77$).

Finally, we looked at just the intervention group's mean scores over time with a one-way repeated measures ANOVA. The results showed that there was no significant change from Time 1 to Time 2 ($Mean_1=62.6$; $Mean_2=65.0$; $F_{1,18}=0.98$; $p=0.34$), but there was a significant change from Time 2 to Time 3 ($Mean_2=65.0$; $Mean_3=74.2$; $F_{1,18}=7.81$; $p \leq 0.05$) (see Figure 2). The associated partial eta^2 effect size for this analysis was also large, 0.31.

To summarize, the significant interaction between group and time can be explained by a significant increase in the intervention group's score on the 10-item sample of the Hay 360 ECI Test over the time before the intervention to the six-month follow up. The control group's scores remained about the same over that period of time. The significant increase in EI for the treatment group was not detected immediately following the intervention; instead, it was observed six months after the EI training.

Table 1. Counts and percentages of 33 emergency medicine resident participants in emotional intelligence training by gender and program year.

	Control	Intervention	Total
Gender			
Female	8 (24.2)	8 (24.2)	16 (48.5)
Male	6 (18.2)	11 (33.3)	17 (51.5)
Total	14 (42.4)	19 (57.6)	33 (100)
Program year			
1	6 (18.2)	6 (18.2)	12 (36.4)
2	4 (12.1)	5 (15.2)	9 (27.3)
3	4 (12.1)	8 (24.2)	12 (36.4)
Total	14 (42.4)	19 (57.6)	33 (100)

Table 2. Analysis of variance (ANOVA) results of emotional intelligence measures of 33 emergency medicine residents comparing intervention group, gender and program year level (PGY) at two time periods: Before Intervention, and six-months later.

Source: 4-Way ANOVA with 1 repeated factor (Time)	F	P value	Effect size*
Main effects			
Group (experimental vs. control)	0.26	NS	
Gender	1.01	NS	
PGY level	1.02	NS	
Time (pre-intervention vs. 6-month follow-up)	7.49	0.012	0.263
2-way interactions			
Gender x time	0.03	NS	
Group x time	7.16	0.014	0.254
PGY level x time	0.33	NS	
Group x gender	1.25	NS	
Group x PGY level	0.13	NS	
Gender x PGY level	0.88	NS	
3-way interactions			
Gender x group x time	0.15	NS	
PGY level x group x time	1.03	NS	
PGY level x gender x time	1.77	NS	
Group x PGY level x gender	0.18	NS	
4-way interactions			
Gender x group x PGY level x time	0.94	NS	

NS, not significant
*Partial Eta Squared

DISCUSSION

We developed an educational intervention designed to improve EI as characterized by the skill of SPT. The intervention was customized for EM residents, and was intended to be brief, so as not to compete for curricular time with other topics. Our review of the literature yielded very few studies on attempts to improve EI in the EM resident population, a medical specialty that commonly works in an emotionally charged environment and is prone to professional burnout.[36,37] Our hope is that this study can serve as impetus for continued efforts to study EI improvement in EM residents.

As expected, pre- and post-EI scores for our control group remained relatively stable. The stability in the control group's scores over time, taken together with the changes

Table 3a. Pre emotional intelligence score. Descriptive statistics (mean, standard deviation [SD], and number) of emotional intelligence measure for 33 emergency medicine residents broken down by gender, post-graduate year (PGY) and intervention/control groups before intervention.

Group	PGY-level	Mean	SD	N
Female				
Intervention	PGY 1	55.0	13.2	3
	PGY 2	70.0	14.1	2
	PGY 3	50.0	17.3	3
	Total	56.9	15.3	8
Control	PGY 1	65.0	18.0	3
	PGY 2	72.5	10.6	2
	PGY 3	65.0	5.0	3
	Total	66.9	11.3	8
Total	PGY 1	60.0	15.2	6
	PGY 2	71.3	10.3	4
	PGY 3	57.5	14.1	6
	Total	61.9	14.0	16
Male				
Intervention	PGY 1	68.3	2.9	3
	PGY 2	63.3	5.8	3
	PGY 3	68.0	7.6	5
	Total	66.8	6.0	11
Control	PGY 1	70.0	17.3	3
	PGY 2	62.5	10.6	2
	PGY 3	65.0	0.0	1
	Total	66.7	12.5	6
Total	PGY 1	69.2	11.1	6
	PGY 2	63.0	6.7	5
	PGY 3	67.5	6.9	6
	Total	66.8	8.5	17
Total				
Intervention	PGY 1	61.7	11.3	6
	PGY 2	66.0	8.9	5
	PGY 3	61.3	14.3	8
	Total	62.6	11.7	19
Control	PGY 1	67.5	16.0	6
	PGY 2	67.5	10.4	4
	PGY 3	65.0	4.1	4
	Total	66.8	11.4	14
Total	PGY 1	64.6	13.6	12
	PGY 2	66.7	9.0	9
	PGY 3	62.5	11.8	12
	Total	64.4	11.6	33

Table 3b. Six-month follow-up emotional intelligence score. Descriptive statistics (mean, standard deviation [SD], and number) of EI measure for 33 emergency medicine residents broken down by gender, post-graduate year (PGY) and intervention/control groups six months after intervention.

Group	PGY-level	Mean	SD	N
Female				
Intervention	PGY 1	66.7	5.8	3
	PGY 2	75.0	14.1	2
	PGY 3	71.7	7.6	3
	Total	70.6	8.2	8
Control	PGY 1	61.7	14.4	3
	PGY 2	72.5	3.5	2
	PGY 3	65.0	5.0	3
	Total	65.6	9.4	8
Total	PGY 1	64.2	10.2	6
	PGY 2	73.8	8.5	4
	PGY 3	68.3	6.8	6
	Total	68.1	8.9	16
Male				
Intervention	PGY 1	85.0	5.0	3
	PGY 2	80.0	5.0	3
	PGY 3	70.0	11.7	5
	Total	76.8	10.6	11
Control	PGY 1	61.7	22.5	3
	PGY 2	75.0	0.0	2
	PGY 3	65.0	0.0	1
	Total	66.7	15.7	6
Total	PGY 1	73.3	19.4	6
	PGY 2	78.0	4.5	5
	PGY 3	69.2	10.7	6
	Total	73.2	13.1	17
Total				
Intervention	PGY 1	75.8	11.1	6
	PGY 2	78.0	8.4	5
	PGY 3	70.6	9.8	8
	Total	74.2	9.9	19
Control	PGY 1	61.7	16.9	6
	PGY 2	73.8	2.5	4
	PGY 3	65.0	4.1	4
	Total	66.1	12.0	14
Total	PGY 1	68.8	15.5	12
	PGY 2	76.1	6.5	9
	PGY 3	68.8	8.6	12
	Total	70.8	11.4	33

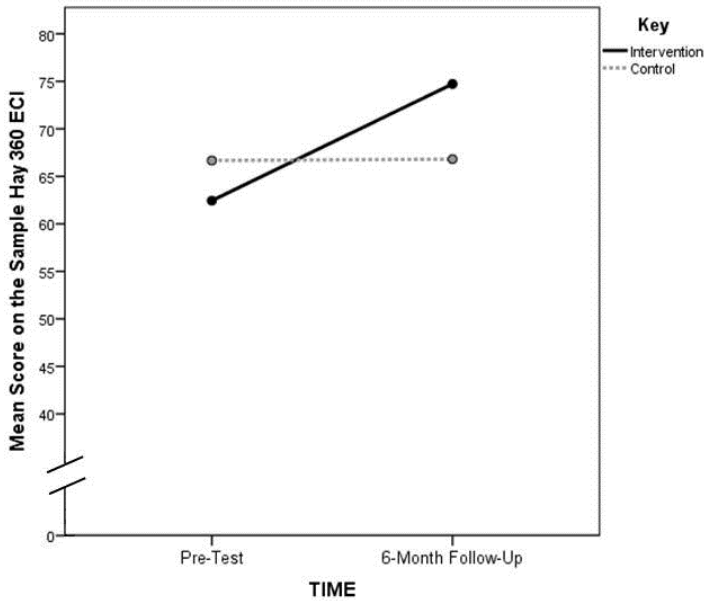


Figure 1. Line graph of the 2-Way Interaction between groups of Emergency Medicine Residents (intervention vs. control) and time (EI pre-test and EI six-month follow up).
ECI, emotional competence inventory

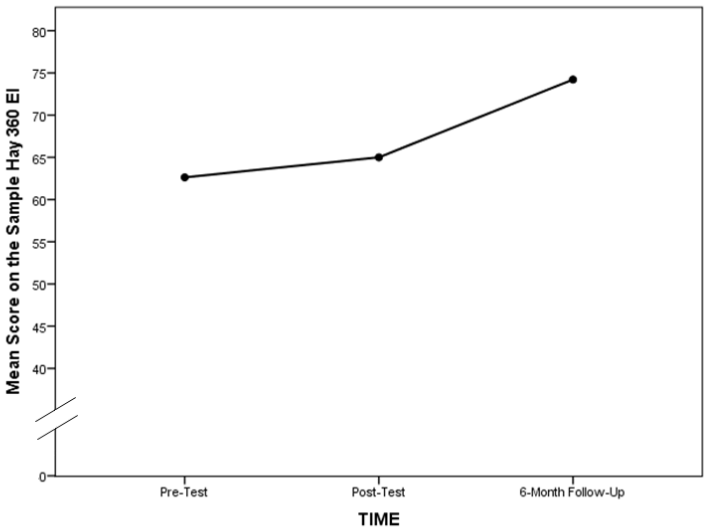


Figure 2. Line graph of the emergency medicine resident intervention group's mean emotional intelligence scores over three time periods: pre-test, post-test, and six-month follow up.
EI, emotional intelligence

in the intervention group's scores indicate that the test was neither intuitive nor learned, so that our results cannot be attributed solely to testing effects. Additionally, the six-month period between our initial and final assessments represented a satisfactory washout period.

We did not detect significant improvement in EI scores immediately following the brief intervention. However, at testing six months post-intervention, the intervention group's EI scores significantly improved, whereas the control group's scores had not. Intuitively, we would have expected the training effect to be apparent at the initial post-intervention assessment. Instead, consistent with other studies, we found that residents required time to process and apply the EI skills training in the clinical setting.[17,22] Perhaps the traditional feedback related to interpersonal communications and received through clinical performance evaluations was more meaningful to the treatment group when compared to the control group.

Prior literature has demonstrated that in the general population, EI improves over a lifetime and is influenced by life experience.[4,11-14] However, in physicians, research has shown that EI tends to decrease over time and particularly during the course of residency training.[11-14] As such, the gains in physician EI we observed after our intervention represent a notable reversal of the decline in EI normally associated with physician development through residency training.

LIMITATIONS

While we tried to study the impact of this brief intervention using an experimental design, there are certainly limitations to this study. First, we believe we did not observe significant main effects across gender or length of training because of the relatively small size of these subgroups within our population. However, we believe that collectively we had sufficient power to accurately assess the effects of the intervention over time. Further studies of the intervention's effect on EM residents at other program sites and to residents in other specialties are needed to determine the generalizability of our findings.

Second, due to time limitations we designed our intervention to specifically focus on one small aspect of EI, the skill of SPT because of its relationship to engagement with others; i.e. you have to notice or observe distress in others in order to be able to act or make a compassionate choice.[32] Yet our small intervention on only one sub-skill of the many that comprise EI was found to improve a score reflective of total EI. Further research is needed to help explain the nature of the relationship between SPT and total EI.

Finally, we acknowledge that a major limitation of this study is that we used a measurement instrument that was only a sample of the full battery of the Hay 360 ECI (or the ECI-U for university students). Further research is needed to determine whether the full battery ECI-U would have yielded the same results.

CONCLUSION

In summary, we were able to observe improved scores on a measure of EI in EM residents with a brief intervention designed to improve the residents' skill in SPT. Interestingly however, the improvement was not immediate and only observed after a six-month delay. We conclude from this finding that the SPT skill required additional practice with specific feedback about interpersonal communications in the clinical

setting to be fully realized by our residents. Those who received the EI training were better able to internalize feedback on social interactions than were trainees without the brief intervention. Given the importance of physician communication skills to patient satisfaction, this research is all the more timely.

Address for Correspondence: Diane L. Gorgas, MD, Ohio State University College of Medicine, 760 Prior Hall, 376 W. 10th Ave., Columbus, OH 43210. Email: diane.gorgas@osumc.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Ciarrochi JV, Chan AY, Caputi P. A Critical Evaluation of the Emotional Intelligence Construct. *Pers Individual Differences.* 2000;28:539-61.
2. Mayer JD, Salovey P, Caruso DR. Emotional Intelligence: Theory, findings, and implications. *Psychol Inq.* 2004;15(3):197-215.
3. Law KS, Wong CS, Song LJ. The construct and criterion validity of emotional intelligence and its potential utility for management studies. *J Appl Psychol.* 2004;89(3):483-96.
4. Galinsky AD, Maddux WW, Gilin D, et al. Why it pays to get inside the head of your opponent" The differential effects of perspective taking and empathy in negotiations. *Psychol Sci.* 2008;19(4):378-84.
5. Goleman D. "Emotional Intelligence." *The Business Voice.* YouTube video. 3 min. 52 sec. Uploaded by: More than Sound Productions on Jan. 26, 2011. Available at: https://youtu.be/wJhfKYzKc0s.
6. Watkin C. Developing Emotional Intelligence. *Int J Select Assess.* 2000;8(2):89–92.
7. Locke EA. Why emotional intelligence is an invalid concept. *J Organ Behav.* 2005;26:425-431.
8. Landy FJ. Some historical and scientific issues related to research on emotional intelligence. *J Organ Behav.* 2005;26:411-424.
9. Kahn J. Can Emotional Intelligence be taught? *New York Times Sunday Magazine*. Sept 15, 2013: MM44.
10. Nelis D, Quoidbach J, Mikolajczak M, et al. Increasing emotional intelligence: (How) is it possible? *Pers Individual Differences.* 2009;47:36-41.
11. Ishak WW, Lederer S, Mandili C, et al. Burnout during residency training: A literature review. *J Grad Med Educ.* 2009;1(2):236-42.
12. Bellini LM and Shea JA. Mood change and empathy decline persist during three years of internal medicine training. *Acad Med.* 2005;80(2):164-7.
13. Lee FJ, Stewart M, Brown JB. Stress, burnout, and strategies for reducing them: What's the situation among Canadian family physicians? *Cam Fam Physician*. 2008;54:234-5.e1-5.
14. Neumann M, Edelhauser F, Tauschel D, et al. Empathy Decline and its reasons: A systematic review of studies with medical students and residents. *Acad Med*. 2011;86(8):996-1009.
15. Satterfield J, Swenson S, Rabow M. Emotional intelligence in internal medicine residents: Educational implications for clinical performance and burnout. *Ann Behav Sci Med Educ*. 2009;14(2):65-8.
16. Webb AR, Young RA, Baumer JG. Emotional Intelligence and the ACGME Competencies. *J Grad Med Educ.* 2010;2(4):508-12.
17. Satterfield JM and Huges E. Emotion skills training for medical students: A systematic review. *Med Educ.* 2007;41:935-41.
18. Mayer JD, DiPaolo M, Salovey P. Perceiving affective content in ambiguous visual stimuli: A component of Emotional Intelligence. *J Pers Assess.* 1990;54(3&4)772-81.
19. Vakola M, Tsaousis I, Nikolaou I. The role of emotional intelligence and personality variables on attitudes toward organizational change. *J Manage Psychol.* 2004;19(2):88-110.
20. Krasner MS, Epstein RM, Beckman H, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. *JAMA.* 2009;302(12):1284-93.
21. Boyatzis RE, Stubbs EC, Taylor SN. Learning cognitive and Emotional Intelligence competencies through graduate management education. *Acad Manage Learn Edu.* 2002;1(2):150-62
22. Stoller JK, Taylor CA, Farver CF. Emotional intelligence competencies provide a developmental curriculum for medical training. *Med Teach.* 2013;35(3):243-7.
23. Cherry MG, Fletcher I, O'Sullivan H, et al. BEME Guide No. 17: What impact do structured educational sessions to increase emotional intelligence have on medical students? *Med Teach.* 2012;34(1):11-19.
24. Arora S, Ashrafian H, Davis R, et al. Emotional intelligence in medicine: A systematic review through the context of the ACGME competencies. *Med Educ.* 2010;44(8):749-64.
25. MacCann C, Matthews G, Zeidner M, et al. Psychological assessment of emotional intelligence: A review of self-report and performance-based testing. The International Journal of Organizational Analysis. 2003;11(3):247-74.
26. Conte JM. A review and critique of emotional intelligence measures. *J Organ Behav.* 2005;26(4):433–40.
27. Wolff SB. *Emotional Competence Inventory (ECI) Technical Manual*. 3rd ed. Philadelphia, PA: The Hay Group, McClelland Center for Research and Innovation; 2006.
28. Dugan JP, Bohle CW, Woelker LR, et al. The role of social perspective-taking in developing students' leadership capacities. *J Student Affairs Research Practice*. 2014;51(1):1–15.
29. Davies M, Stankov L, Roberts RD. Emotional intelligence: In search

of an elusive construct. *J Pers Soc Psychol.* 1998;75:989-1015.

30. Stratton TD, Elam CL, Murphy-Spencer AE, et al. Emotional intelligence and clinical skills: Preliminary results from a comprehensive clinical performance examination. *Acad Med.* 2005;80(10 supp):S34-S37.
31. Goleman D, Boyatzis R, Hay Group Inc. "Hay Group Transforming Learning EI Quiz." Hay Group Inc. Available at: https://atrium.haygroup.com/us/quizzes/emotional-intelligence-quiz.aspx. Accessed Jun 3, 2009.
32. Goleman D. Compassion. *Presented at the TED2007 Conference*. Monterey, CA. 2007. Available at: http://www.ted.com/talks/daniel_golman_on_compassion. Accessed Jan 9, 2012.
33. Brown JD. Statistics Corner. Questions and answers about language testing statistics: Effect size and eta squared. *Shiken: JALT Testing & Evaluation SIG Newsletter*. 2008;12(2):38-43. Available at: http://jalt.org/test/PDF/Brown28.pdf. Accessed Nov 19, 2013.
34. Keppel G. *Design & analysis: A researcher's handbook.* 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc. 1982.
35. Stevens J. *Intermediate statistics: A modern approach*. (p 146) Hillsdale NJ: Lawrence Erlbaum Associates, Publishers. 1990.
36. Cydulka RK and Korte R. Career satisfaction in emergency medicine: The ABEM longitudinal study of emergency physicians. *Ann Emerg Med*. 2008;51(6):714-22.
37. Kimo Takayesu J, Ramoska EA, Clark TR, et al. Factors associated with burnout during emergency medicine residency. *Acad Emerg Med.* 2014;21(9):1031-5.

Correlation of Simulation Examination to Written Test Scores for Advanced Cardiac Life Support Testing: Prospective Cohort Study

Suzanne L. Strom, MD*
Craig L. Anderson, PhD, MPH†
Luanna Yang, PharmD, MD‡
Cecilia Canales, MPH‡
Alpesh Amin, MD, MBA§
Shahram Lotfipour, MD, MPH†
C. Eric McCoy, MD, MPH†
Mark I. Langdorf, MD, MHPE†

*University of California Irvine School of Medicine, Department of Anesthesia and Perioperative Care, Irvine, California
†University of California Irvine School of Medicine, Department of Emergency
‡Medicine, Irvine, California
§University of California Irvine, Irvine, California
University of California Irvine School of Medicine, Department of Medicine, Irvine, California

Section Editor: David Wald, DO
Submission history: Submitted April 23, 2015; Revision received September 18, 2015; Accepted October 19, 2015
Electronically published November 22, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.10.26974

Introduction: Traditional Advanced Cardiac Life Support (ACLS) courses are evaluated using written multiple-choice tests. High-fidelity simulation is a widely used adjunct to didactic content, and has been used in many specialties as a training resource as well as an evaluative tool. There are no data to our knowledge that compare simulation examination scores with written test scores for ACLS courses.

Objective: To compare and correlate a novel high-fidelity simulation-based evaluation with traditional written testing for senior medical students in an ACLS course.

Methods: We performed a prospective cohort study to determine the correlation between simulation-based evaluation and traditional written testing in a medical school simulation center. Students were tested on a standard acute coronary syndrome/ventricular fibrillation cardiac arrest scenario. Our primary outcome measure was correlation of exam results for 19 volunteer fourth-year medical students after a 32-hour ACLS-based Resuscitation Boot Camp course. Our secondary outcome was comparison of simulation-based vs. written outcome scores.

Results: The composite average score on the written evaluation was substantially higher (93.6%) than the simulation performance score (81.3%, absolute difference 12.3%, 95% CI [10.6-14.0%], $p<0.00005$). We found a statistically significant moderate correlation between simulation scenario test performance and traditional written testing (Pearson $r=0.48$, $p=0.04$), validating the new evaluation method.

Conclusion: Simulation-based ACLS evaluation methods correlate with traditional written testing and demonstrate resuscitation knowledge and skills. Simulation may be a more discriminating and challenging testing method, as students scored higher on written evaluation methods compared to simulation. [West J Emerg Med. 2015;16(6):907–912.]

INTRODUCTION

There is early and promising evidence that high-fidelity simulation may be more effective in training healthcare providers in the management of critically ill patients.[1-4]

Previous work has reported its use to assess the psychomotor performance of senior medical students on the American Heart Association's (AHA) standardized Advanced Cardiac Life Support (ACLS) clinical resuscitation scenarios.[5] This research showed that a simulation-based course in ACLS resulted in enhanced student performance, with improved critical action completion, clinical knowledge and psychomotor skill application, and decreased time to cardiopulmonary resuscitation (CPR) and defibrillation.

Student assessment of knowledge acquisition after an ACLS course is traditionally performed using multiple-choice testing alone, with practical skills demonstration of basic airway management, CPR and defibrillation. Although with little evidence to support its use, written evaluations for the assessment of critical management skills has been the historical standard. The advent of evidenced-based medicine and medical simulation has created debate on the optimal evaluation method to assess medical students' ability to manage critically ill patients.

We are not aware of any literature that evaluates the relationship between integrated high-fidelity simulation-based methods and traditional written cognitive testing with non-integrated psychomotor performance.[6] This evaluation was recommended as one of the critical steps of core competency assessment by a professional academic society working group on assessment of observable learner performance.

The objective of our study was to correlate results of a novel high-fidelity simulation-based evaluation method with traditional written evaluation for senior medical students enrolled in an ACLS course.

METHODS

We performed a prospective cohort study evaluating the correlation between high-fidelity simulation-based evaluation with traditional written testing for senior medical students enrolled in an ACLS course. The study was conducted in a medical school simulation center. We obtained institutional review board approval to record simulation sessions and collect patient management data from 19 student volunteers (11 females), most interested in careers in emergency medicine, anesthesiology, or surgery. The course was held over a four-day period in one school week in the last quarter of the senior year. We recorded each student managing a standard acute coronary syndrome (ACS)/ventricular fibrillation (VF) cardiac arrest scenario just prior to the start of the course, and then tested them in both written and simulation format (identical cardiac arrest scenario) upon completion of the course. The 32-hour course consisted of 12 hours of didactics, eight hours of simulation training, eight hours of self-study time, and four hours of post-course practical and written testing.

The three traditional written evaluation instruments were the following: 1) a multiple-choice test, 2) a cardiac rhythm test, and 3) a clinical management test. The 36 questions of multiple choice were developed by the AHA, which covered the breadth of content from the ACLS Student Manual. The questions focused on basic and advanced airway management, algorithm application, resuscitative pharmacology, and special situations like drowning and stroke recognition. The rhythm knowledge evaluation consisted of 20 examples of various brady-and tachyarrhythmias, heart blocks and asystole/agonal rhythm to which the students were required to match rhythm diagnoses on a one-to-one basis. The clinical management "therapeutic modalities" was a fill-in-the-blank test including seven clinical scenarios: ACS, symptomatic bradycardia, pulseless electrical activity, refractory VF, stable and then unstable ventricular tachycardia, third-degree heart block, and asystole (appendices 1, 2 and 3). All written evaluation tools were based on content from the ACLS student manual or obtained from the AHA. All testing protocols and tools were evaluated by two expert ACLS instructors/experienced clinicians (anesthesiologist and emergency physician) prior to implementation of the course. Although we weighted the three components equally in the composite "correct answer" score, the maximum possible written test points were 36 (multiple choice), 20 (rhythms) and 61 ("therapeutic modalities").

To assess post-course ACLS skills, students directed a high-fidelity simulation scenario of a patient with ST-elevation acute myocardial infarction, VF cardiac arrest, defibrillation, basic and advanced airway management, return of spontaneous circulation (ROSC), third-degree heart block, hypotension, acidosis and activation of the cardiac catheterization team. The simulation-based assessment was clinically oriented and approximated the course of events that would take place in the management of a real patient. Each student was tested without additional team members to whom they would normally delegate tasks.

We judged resuscitation successful and awarded ROSC if the student began near-continuous CPR, performed effective bag-valve-mask and/or endotracheal intubation, defibrillated with appropriate joules, and administered two correct doses of epinephrine (or one of vasopressin) and either lidocaine or amiodarone in appropriate doses. We calculated the Kappa statistic for inter-rater reliability. Disagreements in scoring were resolved by jointly reviewing the videos.

Students performed their simulations in a state-of-the-art simulation center approximating a resuscitation room in a modern emergency department. The equipment used in the 65,000+ square-foot medical simulation center included a SimMan 3G © (Laerdal, Wappinger Falls, NY), live defibrillator and crash cart, cardiac monitor, and basic and advanced airway equipment. We used B-line Medical Simbridge software © (B-line Medical, Washington, DC) for video capture, storage and review.

A technical skills checklist of critical actions for the scenarios was created by clinical and simulation faculty using a modified-Delphi technique. Prior to participation in the ACLS course, subjects were recorded performing as team leader in the

standard simulation scenario. The students then completed the Resuscitation Boot Camp with imbedded ACLS course and, as a final test, each student was recorded repeating the same ACS/VF scenario (12-15 minutes). Two expert ACLS instructors (one a regional faculty member) scored the recordings of the before and after performances separately on a 121-point scale and the mean of their assessments was used for analysis. To foster inter-rater reliability, the two instructors jointly developed the scoring scheme, identified each action item, agreed to meaning of the description of the action, and assigned point values. The instructors were not blinded to the study hypothesis, but were blinded to the students' written test performance.

Our primary outcome measure was the correlation between the simulation-based evaluation method and the traditional written evaluation. Our secondary outcome was the comparison of the two scores between the modalities.

We excluded one student who scored very poorly on the written test component of cardiac rhythm interpretation at 55% correct. All other students scored 90-100% on this testing modality. The excluded student's overall score was 78.0% correct, while all other students scored means of 86.6-98.1% correct. Therefore, the excluded student was a clear outlier.

We used t-tests for paired data to compare written and simulation test scores, with each student serving as their own control. We used linear regression to quantify the relationship between the two sets of scores, and set statistical significance at $p<0.05$. (STATA version 12.1, StataCorp, College Station, Texas)

Table. Individual and grouped percent correct performance scores for traditional written evaluation vs. simulation evaluation.

	Mean±SD
Multiple choice test	89.4±5.7%
Cardiac rhythm test	97.8±10.7%
Clinical management test	93.8±6.3%
Mean of three written tests	93.6±5.0%
Simulation test	81.3±3.2%
Difference	12.3±3.5%

SD, standard deviation

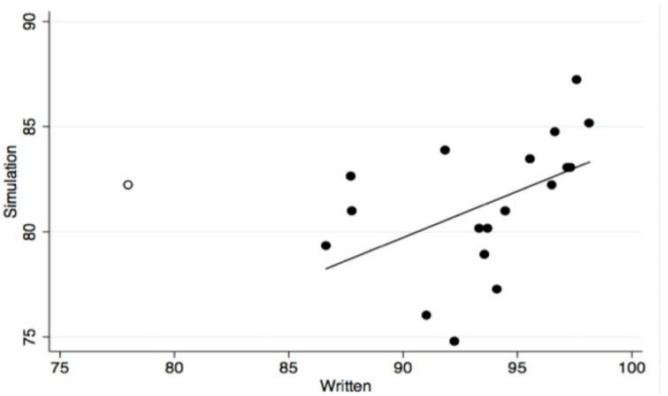


Figure. Correlation between mean percent correct score on traditional three-component written evaluation vs. percent correct score on simulation evaluation. Open circle student was excluded due to outlying low score on cardiac rhythm test.

RESULTS

The composite average score on the three written evaluations was substantially higher (93.6%) than the simulation performance score (81.3%, absolute difference 12.3%, 95% CI [10.6-14.0%], $p<0.00005$). The various component mean and SD scores are listed in the Table.

We found a statistically significant moderate correlation between simulation scenario test performance and traditional written test performance (Figure) (Pearson $r=0.48$, $p=0.04$).

Inter-rater reliability for scoring the participants in pre- and post- training scenarios was good. The median kappa for the 75 test items was 0.68 (interquartile range 0.36-0.94). Forty-six items (61%) had kappa >0.60.

DISCUSSION

We found that high-fidelity simulation-based evaluation and traditional written testing for senior medical students in an ACLS course correlates well with each other. Simulation is being incorporated in the education, training, and evaluation of healthcare providers at a rapid pace. As educational technology advances rapidly, the research to support its use has lagged behind. Traditional written evaluations are widely used, and have been accepted as the standard for healthcare providers' ability to manage critical patients. However, as simulation is realistic, actively engaging and clinically based, healthcare teachers have begun to question written testing.[7]

We found a positive moderate correlation between simulation-based evaluation and traditional written evaluation. Other studies have compared the two but did not specifically assess correlation, nor report results in medical trainees. Rodgers' study on nursing students in an ACLS course completing both a written and practical evaluation concluded that written evaluation is, not surprisingly, a poor predictor of skill performance.[8] Issenberg similarly found no association between CPR psychomotor skills and total knowledge in nursing students.[9] As physician trainees are destined to be team leaders in resuscitation, our work is the first to study medical students, and therefore adds to this literature.[10,11]

The issue of correlation should not be misconstrued as equivalence. We contend that the simulation evaluation is superior to evaluate psychomotor skills, yet accept the place of written evaluation to demonstrate cognitive mastery across broad medical content. The correlation demonstrates that students who have traditionally done well in written testing are likely to also do well in a simulation evaluation. An educator should acknowledge that the two evaluation methods are complementary, rather than substitutable, and consider adding such evaluation to tasks that require manual dexterity and critical thinking. Furthermore, simulation requires substantial human and capital resources to show competence,

and therefore limits its widespread application. Simulation, by necessity, focuses on narrow clinical scenarios, which, though chosen to represent critical management, cannot cover the entire breadth of cardiac resuscitation. Hence both simulation and written evaluations are likely necessary.

Our secondary outcome compared the two scores. The composite average score on the written evaluation was substantially higher (93.6%) than the simulation performance score (81.3%). It is important to note that all 19 students had the same training and were evaluated by both written and simulation methods. A higher written test score does not mean better performance, as the two modalities measure different outcomes.

Participants find high-fidelity simulation for critical event management to be a valuable educational experience.[12] Emotional arousal is effective in memory acquisition[13] and simulation-based experiential learning has been shown to be effective in retention of skills,[14] improving clinical outcomes,[15] and reducing error related healthcare costs.[16] Furthermore, repetition of simulation experience reinforces knowledge acquisition and increases confidence.[17,18]

Written testing has historically been the most common mode of evaluation. However, the construct validity of the AHA's ACLS test has been challenged, as nurses' scores were not shown to correlate with performance on resuscitation after an ACLS course. These same authors opined that the written testing at least had content validity, as the tests questions were drawn directly from the student manual. Finally, their analysis supported our contention that the two modalities complement each other in providing a broad assessment of the learner's performance.[19]

Despite these questions, newer examination techniques, such as simulation-based evaluations, need to be validated before widespread use. Our study provides preliminary evidence that will shape this discussion. There has been a move toward simulation for assessment, as exemplified by a report of five years of certification via Fundamentals of Laparoscopic Surgery.[20] In addition, simulation is used in both initial and maintenance of certification in anesthesiology.[21-23] Hence, it is critical to scrutinize new testing methods to validate that they at least approximate traditional techniques.

Students performed better on the written form of testing than on the simulation. We believe this indicates that the simulation evaluation method is a more demanding measure, which emphasizes application of knowledge over rote memorization. Furthermore, we found a narrower range of student performance with the simulation method (range of scores 74.8-87.2%, Δ12.4%) than the written assessment (80.9-98.2%, Δ17.3%), which indicates a more uniform and direct performance in concert with course goals. Since the purpose of the boot camp is to prepare students for clinical practice, an instrument/method that better generates a consistent execution of skills is valued over abstract knowledge applied in isolation. In the end, the educator should consider using both methods of evaluation when teaching psychomotor skills. In addition, a quantitative simulation evaluation with an established "pass" threshold should be incorporated, in order to move toward competency-based evaluation.

LIMITATIONS

Our study has limitations, including enrolling a small sample of self-selected, highly motivated students entering the fields of emergency medicine, anesthesiology, or surgery. We did not have any baseline data on the subjects' prior ACLS training or experience. However, this did not affect our study's ability to evaluate the relationship between a simulation-based and written evaluation tool, as students served as their own controls. We used a previously non-validated simulation evaluation scale with arbitrary weighting of points for critical actions (derived from two expert ACLS instructors), though the action items had been used for grading in the course for 15 years. Furthermore, our assessment tool is based on AHA guidelines, and is clinically focused on critical action items that approximate real clinical care, compared to a multiple-choice or even fill-in-the-blank format. Our course was non-traditional and expanded from ACLS, and included advanced airway management and additional didactics. However, both assessment methods tested knowledge and skills from this non-traditional course format, which would not confound the assessment methods themselves. Our criterion reference was the ACLS written exam. To our knowledge, these test questions are not analyzed for reliability or validity. There are no previous studies that demonstrate construct validity of the AHA written examination or correlate clinical performance with the written examination. The correlation between written and simulation examination performance in this study does demonstrate some degree of construct validity. The written examination is based entirely on the ACLS manual and should therefore have content validity.

To provide maximal experience with simulation and to reinforce specific and detailed proper ACS/cardiac arrest management, we used the same teaching and testing scenario and informed the students that the pre- and post-tests would be identical. This may have artificially improved post-test performance through studying specifically for the known test, as well as additional familiarity with the simulation technology. We did not control for progressive experience and therefore comfort with the mannequin or simulation experience, nor was there a traditional ACLS course student control group.

Future studies should use students destined for all specialty residencies, and assess the rate of long-term retention of psychomotor skills.

We excluded one outlier who scored far below the other students, at 55% correct on the rhythm matching test (11/20 correct). This student scored average on the simulation evaluation, which only required identification of three (not 20) obvious rhythms. Including this outlier would have made our correlation fall short of statistical significance. However, the scatter plot visually demonstrates the conclusion that higher

written scores are associated with higher simulation scores. With our small sample size, one outlier has a higher possibility of skewing results away from statistically significant correlation. Further research will be needed to determine if exclusion of this outlier was appropriate.

We did not study, nor do we advocate, any particular "pass" threshold for simulation evaluation. As in any other course, the instructor would need to establish this given the difficulty of content, ability of students to master material with the course format, and degree of "high stakes" activity.

The three components of the written testing have not been correlated with each other, as they are designed to test different cognitive skills. Therefore, correlation of their aggregate with simulation evaluation may lack a basic level of validation. Nevertheless, the simulation is new, labor intensive and expensive, and therefore more in need of scrutiny and validation. Our results of testing relatively novice learners may not be generalizable to more experienced providers. Lastly, the simulation evaluation raters had, at best, vague familiarity with the students. That they were identifiable on the recordings may have introduced an unknown bias into the evaluation.

CONCLUSION

This study is the first to compare written and simulation-based evaluation in medical students. Simulation-based ACLS evaluation methods correlate with traditional written evaluation methods, and provide additional opportunity to demonstrate competency of resuscitation knowledge and skills.

Simulation may be a more discriminating and challenging testing method, as students scored higher on written evaluation methods compared to simulation. The meaning of this difference needs clarification through further research.

Address for Correspondence: Mark I. Langdorf, MD, University of California Irvine School of Medicine, Rt 128, 101 The City Drive, Orange, CA 92868. Email: milangdo@uci.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. McCoy CE, Menchine M, Anderson C, et al. Prospective randomized crossover study of simulation vs. didactics for teaching medical students the assessment and management of critically ill patients. *J Emerg Med*. 2011;40(4):448-55.
2. Wang CL, Schopp JG, Petscavage JM, et al. Prospective randomized comparison of standard didactic lecture versus high-fidelity simulation for radiology resident contrast reaction management training. *AJR AM J Roentgenol*. 2011;196(6):1288-95.
3. Wayne DB, Butter J, Siddal VJ, et al. Simulation-based training of internal medicine residents in advanced cardiac life support protocols: a randomized trial. *Teach Learn Med*. 2005;17(3):210-6.
4. Steadman RH, Coates WC, Huang YM, et al. Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. *Crit Care Med*. 2006;34:151-7.
5. Landgorf MI, Strom SL, Yang L, et al. High-fidelity simulation enhances ACLS training. *Teach Learn Med*. 2014;26(3):266-73.
6. Kessler CS and Leone KA. "The current state of core competency assessment in emergency medicine and a future research agenda: recommendations of the working group on assessment of observable learner performance." *Acad Emerg Med*. 2012;19(12):1354-9.
7. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387-96.
8. Rodgers DL, Bhanji F, McCkee BR. Written evaluation is not a predictor for skills performance in an Advanced Cardiovascular Life Support course. *Resuscitation*. 2010;81(4):435-6.
9. Roh YS and Issenberg SB. Association of cardiopulmonary resuscitation psychomotor skill with knowledge and self-efficacy in nursing students. *Int J Nurs Pract*. 2014;20(6):674-9.
10. Dal U and Sarpkaya D. Knowledge and psychomotor skill of nursing students in North Cyprus in the area of cardiopulmonary resuscitation. *Pak J Med Sci*. 2013;29(4):966-71.
11. Madden C. Undergraduate nursing students' acquisition and retention of CPR knowledge and skills. *Nurse Educ Today*. 2006;26(3):218-27.
12. Lighthall GK, Barr J, Howard SK, et al. Use of a fully simulated intensive care unit environment for critical event management training for internal medicine residents. *Crit Care Med*. 2003;31(10):2437-43.
13. Cahill L and McGaugh JL. Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*. 1998;21(7):294–99.
14. Wayne DB, Siddall VJ, Butter J, et al. A Longitudinal Study of Internal Medicine Residents' Retention of Advanced Cardiac Life Support Skills. *Acad Med*. 2006;81(10):S9-12.
15. Draycott T, Sibanda T, Owen L, et al. Does training in obstetric emergencies improve neonatal outcome? *BJOG*. 2006;113(2):177-82.
16. Cohen ER, Feinglass J, Barsuk JH, et al. Cost savings from reduced catheter-related bloodstream infection after simulation-based education for residents in a medical intensive care unit. *Simulation in Healthcare*. 2010;5(2):98–102.
17. O'Brien G, Haughton A, Flanagan B. Interns' perceptions of performance and confidence in participating in and managing simulated and real cardiac arrest situations. *Med Teach*.

2001;23(4):389-95.

18. Jung J, Shilkofski N, Perretta J, et al. A brief simulation-based curriculum for medical students improves performance of essential resuscitation skills. *Sim Health*. 2010;5(6):397.
19. Rodgers DL, Bhanji F, McKee BR. Written evaluation is not a predictor for skills performance in an Advanced Cardiovascular Life Support course. *Resuscitation*. 2010;81(4):453-6.
20. Okrainec A, Soper NJ, Lee L, et al. Trends and results of the first 5 years of Fundamentals of Laparoscopic Surgery (FLS) certification testing. *Surg Endosc*. 2011;25:1192-98.
21. McIvor W, Burden A, Weinger MB, et al. Simulation for Maintenance of Certification in Anesthesiology: The First Two Years. *J Contin Educ Health Prof*. 2012;32(4):236-42.
22. Miller SH. American Board of Medical Specialties and repositioning for excellence in lifelong learning: Maintenance of Certification. *J Contin Educ Health Prof*. 2005;25(3):151–6.
23. Levine A, Schwartz A, Bryson E, et al. Role of Simulation in US Physician Licensure and Certification. *Mt Sinai J Med*. 2012;79(1):140-53.

How Does Emergency Department Crowding Affect Medical Student Test Scores and Clerkship Evaluations?

Grant Wei, MD*
Rajiv Arya, MD*
Z. Trevor Ritz, DO*
Albert S. He, MD, PhD*
Pamela A. Ohman-Strickland, PhD†
Jonathan V. McCoy, MD*

*Rutgers Robert Wood Johnson Medical School, Department of Emergency Medicine, New Brunswick, New Jersey
†Rutgers School of Public Health, Department of Biostatistics, New Brunswick, New Jersey

Section Editor: Jonathan Fisher, MD
Submission history: Submitted April 29, 2015; Revision received September 25, 2015; Accepted October 19, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.10.27242

Introduction: The effect of emergency department (ED) crowding has been recognized as a concern for more than 20 years; its effect on productivity, medical errors, and patient satisfaction has been studied extensively. Little research has reviewed the effect of ED crowding on medical education. Prior studies that have considered this effect have shown no correlation between ED crowding and resident perception of quality of medical education.

Objective: To determine whether ED crowding, as measured by the National ED Overcrowding Scale (NEDOCS) score, has a quantifiable effect on medical student objective and subjective experiences during emergency medicine (EM) clerkship rotations.

Methods: We collected end-of-rotation examinations and medical student evaluations for 21 EM rotation blocks between July 2010 and May 2012, with a total of 211 students. NEDOCS scores were calculated for each corresponding period. Weighted regression analyses examined the correlation between components of the medical student evaluation, student test scores, and the NEDOCS score for each period.

Results: When all 21 rotations are included in the analysis, NEDOCS scores showed a negative correlation with medical student tests scores (regression coefficient= -0.16, p=0.04) and three elements of the rotation evaluation (attending teaching, communication, and systems-based practice; p<0.05). We excluded an outlying NEDOCS score from the analysis and obtained similar results. When the data were controlled for effect of month of the year, only student test score remained significantly correlated with NEDOCS score (p=0.011). No part of the medical student rotation evaluation attained significant correlation with the NEDOCS score (p≥0.34 in all cases).

Conclusion: ED overcrowding does demonstrate a small but negative association with medical student performance on end-of-rotation examinations. Additional studies are recommended to further evaluate this effect. [West J Emerg Med. 2015;16(6):913–918.]

INTRODUCTION

Emergency department (ED) crowding has been described in emergency medicine (EM) literature as a concern for over 20 years.[1] Previous reports have noted crowding as a risk factor for patients leaving without being seen, increased inpatient mortality, increased frequency of medical errors, and increased

length of stay for all patients.[2-6] Solutions have been proposed with the implementation of surge protocols and improvement of both ED workflow and downstream factors.[7-8] In the academic medical setting, an additional concern is the impact of ED crowding on resident and medical student education.

Relatively few studies exist in the medical literature assessing the effect of ED crowding on the educational outcomes of residents and medical students. Investigations that have been completed do not generally find significant associations between crowding and educational quality. Pines et al. found no association between crowding metrics and resident/medical student assessment of attending physician teaching.[9] Mahler did find that during overcrowded periods, residents saw fewer patients and performed fewer procedures; however, resident physicians judged the quality of their education as unaffected.[10] Perceptions of crowding did not correlate with the perception of educational quality in another study.[11]

The reason for these outcomes may not be immediately obvious, as an intuitive understanding of ED crowding suggests that limitations on space, resources, and attending physician time in periods of crowding would negatively impact secondary goals such as education. It has been speculated that attending physicians may simply prioritize education regardless of crowding status or that crowding status itself may not necessarily impose additional workload on the ED attending physician.[9] Alternatively, less severe ED crowding may improve resident education by increasing the opportunity for residents to see higher volumes and higher acuity patients.[12] Evidentiary support for these ideas is limited.

Several measures of ED crowding have been described previously in the literature. The ED Work Index (EDWIN) identifies ED crowding according to a conceptual formula, with good accordance with physician and nurse impressions of crowding.[13] The National ED Overcrowding Scale (NEDOCS) uses five operational variables in a logistic regression model to identify periods of crowding.[14] Agreement between the models is high and both have good discrimination for prediction of ED crowding.[15]

Given the lack of data in this area, we sought to determine whether ED crowding, as measured by the NEDOCS score, has a quantifiable effect on medical student objective and subjective experiences during the EM clerkship rotations.

METHODS

Study Setting and Population

Our setting is an urban Level It rauma center with ED residents, a required fourth-year medical student clerkship, a 19-member academic faculty, and an approximate annual volume of 70,000 adult patients. The facility contains a separate dedicated pediatric ED as well; data for this study pertained exclusively to the adult ED. The medical student rotation at this facility is scheduled in four-week blocks occurring between July and April with a single combined eight-week block from December-January. Between four and 21 students rotate through the adult ED at a time, completing nine-hour shifts that fall predominately between 9 a.m. and 11 p.m. Medical students are paired up individually with an attending for their shifts, and while there are often other learners (i.e., residents) with the attending there is never more than one clerkship student with an attending.

Medical students complete a "home-grown" exam of objective knowledge at the completion of their rotation and additionally complete a survey regarding their experiences in several core competency areas as specified by the ACGME.[16] The medical student survey evaluates curriculum organization, patient care experience, bedside education, student perception of faculty and residents as educators, problem-based learning and improvement (PBLI), communication, professionalism, and use of systems-based practice on a five-point Likert scale. The student survey is administered to our students at the end of every clinical clerkship, and has been unmodified for several years, providing a readily available anonymous data set that the students were accustomed to provide multiple times a year. It is written and distributed by our Office of Student Affairs, and in addition to providing data for the ACGME, is used to generate anonymous feedback and highlight areas of excellence and deficiency for each clerkship.

Study Protocol

Approval was obtained from this institution's research review board prior to the initiation of any data collection. Between July 2010 and May 2012, medical student survey results, end-of-rotation test scores, and NEDOCS scores were collected and reviewed on an ongoing basis following the completion of each rotation. This period included 20 four-week rotation blocks and one eight-week rotation block for a total of 21 blocks. A total of 211 students rotated through the ED in this time period, with between five and 21 (in the eight-week block) students rotating in the department at one time.

Measurements

First described in 2004, the NEDOCS score quantifies the level of ED crowding by measurement of several variables related to current ED patient load, admitted patient boarding time, and available hospital beds.[14] The resulting score is divided into ranges of values denoting normal volume (NEDOCS<60), busy (61-100), crowded (101-140), dangerously crowded (141-180), and disaster-level crowding (>180). These results appear in Table 1. Data were obtained for this via review of records from the ED's electronic medical record (EDIMS LLC; Parsippany, NJ). The NEDOCS score is calculated and recorded hourly at our institution, and the crowding numbers from 9 a.m. to 11 p.m. of each block were analyzed.

Student test scores had a maximum value of 100 points; student survey responses were scored on a five-point Likert scale, where one indicated strong disagreement with a given statement and five indicated strong agreement. The student

survey appears in Appendix 1.

Data Analysis

We calculated means for student survey results and end-of-rotation test score by rotation period. Weighted regression analyses determined the association between either average medical student test scores or average survey responses and percent crowding among the 21 blocks between the hours of 9 a.m. to 11 p.m. Weighting simultaneously accounted for heterogeneous variation in the block-level means due to differences between periods in the standard deviations (SDs) (in some cases the largest SD was 10 times as large as the smallest SD) and in the number of students (ranging from five to 21 for test scores and from four to 14 for the survey). The weights were set equal to the inverse of the squared standard errors of the block average for each particular response. Regression coefficients, p-values and r-squared values were calculated to assess the fit of the weighted regression models. We performed all data analysis using SAS 9.3 for Windows.

RESULTS

During the period of the study, 211 students rotated through the ED in 20 four-week blocks and one eight-week block, and a total of 10,047 clinical exposure hours were recorded. Dividing the clinical hours via NEDOCS, 19% of the period recorded as crowded, in 3.4% of the study hours the ED was dangerously crowded; and in 0.7% of the study period, the ED was at disaster-level crowding. These results appear in Table 1. The means, SDs, standard errors and ranges of these for the student survey responses and test scores, based on the 21 blocks of students, are summarized in Table 2. The ranges of standard errors for the means indicate the need for weighted regression analysis. NEDOCS scores were most highly correlated with end-of-rotation test scores ($p=0.0003$) and student evaluations of communication, systems-based practice, and bedside education ($p=0.0059$, 0.023, and 0.016, respectively). In all cases, the association was negative, indicating an inverse correlation between crowding and positive survey responses/end-of-rotation test scores; i.e, the more crowded the ED was, the worse the survey responses and test scores were. However, associations with student evaluations of patient care, faculty/resident teaching, PBLI, and professionalism did not reach statistical significance.

Examination of the NEDOCS score distributions revealed a single outlier that was more than 10 points higher than the next highest block's NEDOCS score. This had the potential to unduly influence the results of the regression analysis; thus,

Table 1. National Emergency Department Overcrowding (NEDOC) scale data for study period.

Block of study period	Average NEDOC score	Percent of time overcrowded	Percent of time dangerously overcrowded	Percent of time at disaster level
1	69.1	16.6%	7.5%	4.8%
2	58.2	8.3%	1.8%	0.9%
3	63.4	13.9%	3.6%	2.2%
4	61.5	12.5%	4.9%	2.0%
5	66.2	16.1%	4.7%	0.2%
6	64.7	15.2%	0.9%	0.0%
7	75.7	25.4%	9.8%	2.0%
8	68.7	18.8%	2.9%	0.7%
9	65.8	17.9%	1.6%	0.0%
10	67.2	16.1%	2.2%	0.0%
11	52.8	6.3%	0.0%	0.0%
12	58.7	8.3%	0.4%	0.0%
13	58.6	7.4%	3.6%	2.2%
14	76.8	27.7%	3.3%	0.0%
15	70.5	20.1%	1.1%	0.0%
16	76.1	28.6%	4.5%	0.7%
17	69.3	19.7%	3.2%	0.3%
18	70.7	21.9%	0.4%	0.0%
19	77.2	30.4%	2.5%	0.0%
20	74.8	25.9%	3.6%	0.4%
21	90.1	41.3%	9.2%	0.0%
Average	68.4	19%	3.4%	0.7%

Table 2. Summary statistics for National Emergency Department Overcrowding Scale (NEDOCS) scores and medical student outcomes along with regression coefficients, p-values, and r-squared values summarizing relationship between NEDOCS scores and the outcomes.

Variable	Mean (SD)	Range of averages	Range of SE*	Regression coefficient	p-value	R^2 value
NEDOCS	68.40 (8.12)	52.8-90.1	--	--	--	--
Test score	83.78 (1.88)	80-86.77	1.04-3.34	-0.16	0.0003	50%
Organization	4.35 (0.24)	3.67-4.68	0.02-0.20	0.0077	0.065	17%
Patient care	3.91 (0.23)	3.34-4.31	0.17-0.80	0.0029	0.66	1%
Bedside education	3.84 (0.34)	3.14-4.54	0.02-0.24	-0.018	0.016	27%
Faculty/resident teaching	4.10 (0.25)	3.60-4.57	0.02-0.19	0.0061	0.20	9%
PBLI	4.04 (0.28)	3.34-4.57	0.00-0.18	0.0069	0.24	7%
Communication	4.19 (0.23)	3.77-4.69	0.02-0.35	-0.013	0.0059	34%
Professionalism	4.26 (0.22)	3.70-4.71	0.00-0.25	0.0055	0.43	4%
Systems-based practice	4.16 (0.23)	3.74-4.61	0.02-0.19	-0.011	0.023	24%

SD, standard deviation; *PBLI,* problem-based learning and improvement
*Range of standard errors (SE) of responses as calculated by block. The square of these values are used for weighting in the weighted regression analysis.

the analyses were repeated with this value excluded. With this value excluded, we recalculated the regression coefficients and p-values (i.e., based on 20 blocks of data). These results appear in Table 3. Similar associations were noted as in Table 2. Again, NEDOCS score demonstrated a negative correlation with test score and student evaluation of bedside education, communication, and systems-based practice.

Further sensitivity analysis (data not shown) examined the effect of NEDOCS score after accounting for month of the year; in this analysis, the effect of the NEDOCS score was largely eliminated. Only end-of-rotation test score remained significant ($p=0.011$) and only when including all 21 blocks (i.e., including the time period with the outlying NEDOCS score). As before, there was an inverse correlation between the end-of-rotation test score and NEDOCS score. P-values for all other parameters were non-significant ($p\geq0.34$). On review of the exam scores from block to block, there is a small tendency towards higher scores at the beginning and the end of the year, with the lowest scores tending to cluster around the middle of the year.

DISCUSSION

Relatively few studies have assessed the relationship between ED crowding and educational outcomes. No previous studies to date have examined quantitative markers of educational performance in this setting. This study demonstrated that higher rates of crowding as measured by the NEDOCS score did have a negative effect on certain aspects of the medical student educational experience; this result is at odds with prior studies showing no relationship between educational measures and ED crowding.[9-11] Prior studies have focused predominantly on resident physicians, making this study unique in its focus.

Certain aspects of the medical student rotation evaluation did achieve a statistically significant correlation with the NEDOCS crowding metric, while others did not. One potential explanation is that those elements that do show a correlation may be those most likely to be affected by a crowded ED (in particular, communication and bedside education). Certain other aspects (e.g., professionalism and patient care) may be relatively unaffected as attending physicians view these as more critical elements to maintain regardless of crowding status.

Attending physicians may employ trade-offs that sacrifice certain aspects of the educational process—bedside teaching, exploring a student's differential, or expanding upon teaching points for example—in favor of retaining strategies that maximize patient flow and direct patient care when the ED is crowded.[17] A previous study of resident evaluation of attending physician teaching quality found little association between attending workload and quality of teaching; other factors (interpersonal skill, willingness to teach) had the greatest effect on perceived teaching quality.[18] Supervision of EM residents is known to be adversely affected by ED crowding;[19] a similar situation may apply to medical student education. There may even be a larger effect as attendings have less individually vested in a student doing a four-week rotation vs. a full EM residency, and EM residents should have more experience learning in the opportunistic, unscheduled, and possibly chaotic learning environment of a crowded department.

The findings demonstrated here conflict somewhat with those of Berger et al., who found no correlation between ED attending physician productivity as measured in relative value units (RVU) and medical student evaluation of their clinical teaching.[20] This study did not address physician productivity per se, but it does raise the issue of whether teaching quality is the primary driver in the medical student performance outcomes. It may be the case that despite unchanged teaching quality by attending physicians, the retained knowledge by students may be lower in a crowded ED setting. If true, this argues that environmental effects of a crowded ED (e.g.,

Table 3. Summary statistics for National Emergency Department Overcrowding Scale (NEDOCS) scores and medical student outcomes when data for single outlying block with NEDOCS >80 were excluded.

Variable	Regression coefficient	p-value	R^2 value
Test score	-0.1400	0.0042	37%
Organization	0.0041	0.43	3%
Patient care	-0.0097	0.23	8%
Bedside education	-0.0260	0.0015	44%
Faculty/resident teaching	-0.0084	0.26	7%
PBLI	-0.0030	0.73	1%
Communication	-0.020	<0.0001	62%
Professionalism	0.0039	0.67	1%
Systems-based practice	-0.0140	0.0024	41%

PBLI, problem-based learning and improvement

noise, lack of workspace, frequent task-switching) may play a greater role in the negative effects on medical student experience than interaction with the attending physicians.

Regardless of the individual medical student rotation evaluation results, the end-of-rotation examination did show a negative correlation with ED crowding. The medical students rotating at the study facility receive mandatory weekly educational lectures as well as a suture workshop during their rotation. These didactic experiences are essentially the same for each block of students (the same PowerPoint presentations are given by different members of our attending faculty), making their ED experience the most variable part of the rotation itself and, presumably, the factor most likely to explain variation in their end-of-rotation test scores. Individual medical student motivations may play a role as well; this is discussed in more detail below.

Future investigations should include similar objective measures of student performance and may benefit from comparisons between multiple measurements of ED crowding (NEDOCS, EDWIN, etc.). Additionally, repeating this study with a cohort analysis of EM-applicant medical students versus non-EM applicant medical students would be of interest; this could better elicit the effect of student motivation on rotation experience A standardized National Board of Medical Examiners (NBME) subject examination has replaced the home-grown examination previously used at the study institution; the study could be repeated once sufficient data has been accumulated to allow for the opportunity to expand this analysis beyond the study institution and to standardize results between different institutions.

LIMITATIONS

Our data indicate a possible confounding effect with month of the year on the association between the NEDOCS score and end-of-rotation exam score. Availability of only two years' worth of data limits the power to analyze this effect fully and may lead to over-adjustment by month. However, some kind of temporal association is not unexpected, as the cohort of medical students rotating in the ED varies in its characteristics throughout the year. In the late summer and fall, medical students who plan to apply for EM residences complete their ED rotations. This likely represents a different group with distinct motivations from those who complete ED rotations later in the year, following the residency application period. The former group is likely more aggressive in attaining educational goals despite the potentially crowded state of the ED. Presumably, they may also have had exposure to an ED in the past, giving them some familiarity of how to function in a busy environment. Additionally the number of medical students in a block has significant variation–during the study period we combined what was previously the end of November through January blocks into one extended block and accepted more students into that block. Immediately after the change we reviewed both the student exam scores and their feedback from having a more spread-out schedule and found no differences from the other blocks and previous years. Based on this, we did not specifically exclude the data from that block, This explanation remains speculative, however, and requires further investigation.

Limitations of this study also include small and differing sample sizes within each group of medical students. As noted previously, there is a suggestion of a confounding effect due to month of the year, but more data would be needed to confidently estimate its effect. The study was limited to a single site and did not include the full 24 hours of ED crowding data (We analyzed only the periods of 9 a.m. to 11 p.m. as students do only one overnight shift during their rotation, and the NEDOCS score often falls dramatically during the overnight hours.) When we designed the study, we were aware of how seldom we were crowded on the overnight hours, and that combined with the facts that the students did few overnight shifts led us to exclude the overnight shift in the initial design of the study. In retrospect we probably should have left the overnight shift in the data set and used their exclusion/inclusion as another variable. We analyzed the medical student performance on the basis of their averaged group performance, rather than individual student results for each block in question. Additionally, the end-of-rotation examination used in this study was developed jointly by the clerkship directors and educational faculty within the study site and our sister institution; it is not a standardized examination, and it is not validated. Our site has begun using the NBME advanced subject exam in EM, but it was not yet available at the time of this study. Also, the medical student experience survey, which is administered by our medical school at the end of every clinical clerkship, is to the best of our knowledge also a home-grown and not validated survey. It was written or at least modified by members of the Office of Student Affairs to capture data for the ACGME and within institution use. It was selected as a marker of subjective

student experience because the students, and through them the dean and department chairs, at our school use this survey as the primary method to obtain student feedback on their clinical clerkships. Individual medical student motivations also lie outside the ability of this study to detect, though a cohort analysis of EM-applicant medical students versus non-EM applicants could potentially isolate the effect of medical student motivation as a contributor to outcome.

CONCLUSION

Our study sought to assess connections between objective measures of ED crowding and objective and subjective measures of medical student experience in the ED. A weak negative association was noted between end-of-rotation test scores and NEDOCS scores when considering the entire time period of this study and accounting for variability associated with month of the year. No subjective measure of rotation experience was correlated with the NEDOCS score when accounting for month of the year, which is in accordance with prior studies that have not suggested any effect of ED crowding on medical education.[9,10] The limited association found in this study suggests that ED crowding has a negative effect on medical student education. These results can be applied practically now to schedule medical student rotations for periods in which crowding is expected to be lower, potentially leaving open the opportunity for more educational time.

Address for Correspondence: Grant Wei, MD, Rutgers Robert Wood Johnson Medical School, One Robert Wood Johnson Place MEB Rm. 104, New Brunswick, NJ 08901. E-mail: weigr@rwjms.rutgers.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Dickinson G. Emergency department overcrowding. *CMAJ*. 1989;140:270-1.
2. Weiss SJ, Ernest AA, Derlet R, et al. Relationship between the National ED Overcrowding Scale and the number of patients who leaves without being seen in an academic ED. *Am J Emerg Med.* 2005;23:288-94.
3. Bernstein SL, Aronsky D, Duseja R, et al. The effect of emergency department crowding on clinically oriented outcomes. *Acad Emerg Med*. 2008;16:1-10.
4. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Med J Aust*. 2006;184(5):213-6.
5. Epstein SK, Huckins DS, Liu SW, et al. Emergency department overcrowding and risk of preventable medical errors. *Intern Emerg Med.* 2012;7:173-80.
6. McCarthy ML, Zeger SL, Ding R, et al. Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients. *Ann Emerg Med.* 2009;54:492-503.
7. Moseley MG, Dickerson CL, Kasey J, et al. Surge: an organizational response to emergency department overcrowding. *J Clin Outcomes Manage*. 2010;17(10):453-7.
8. Schiff G. System dynamics and dysfunctionalities: levers for overcoming emergency department overcrowding. *Acad Emerg Med.* 2011;18:1255-61.
9. Pines JM, Prabhu A, McCusker CM, et al. The effect of ED crowding on education. *Am J Emerg Med*. 2010;28:217-20.
10. Mahler SA, McCartney JR, Swoboda TK, et al. The impact of emergency department overcrowding on resident education. *J Emerg Med.* 2012;42(1):69-73.
11. Hoxhaj S, Moseley MG, Fisher A. Resident education does not correlate with the degree of emergency department overcrowding. *Ann Emerg Med.* 2004;44(4):S7.
12. Shayne P, Lin M, Ufberg J. The effect of emergency department crowding on education: blessing or curse? *Acad Emerg Med.* 2009;16:76-82.
13. Bernstein SL, Verghese V, Leung W, et al. Development and validation of a new index to measure emergency department crowding. *Acad Emeg Med.* 2003;10:938-42.
14. Weiss SJ, Derlet R, Arndahl J, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the national ED overcrowding study (NEDOCS). *Acad Emerg Med.* 2004;11:38-50.
15. Weiss SJ, Ernst AA, Nick TG. Comparison of the National Emergency Department Overcrowding Scale and the Emergency Department Work Index for quantifying emergency department crowding. *Acad Emerg Med*. 2006;13:513-8.
16. Nasca T, Philibert I, Brigham T, et al. The next GME accreditation system—rationale and benefits. *N Engl J Med.* 2012;366:1051-6.
17. Aldeen AZ and Gisondi MA. Bedside teaching in the emergency department. *Acad Emerg Med.* 2010;13:460-6.
18. Kelly SP, Shaprio N, Woodruff M, et al. The effects of clinical workload on teaching in the emergency department. *Acad Emerg Med.* 2007;14:526-31.
19. Jelinek GA, Weiland TJ, Mackinlay C. Supervision and feedback for junior medical staff in Australian emergency departments: findings from the emergency medicine capacity assessment study. *BMC Med Educ.* 2010;10:74.
20. Berger TJ, Ander DS, Terell ML, et al. The impact of the demand for clinical productivity on student teaching in academic emergency departments. *Acad Emerg Med.* 2004;11:1364-67.

Medical Student Performance on the National Board of Medical Examiners Emergency Medicine Advanced Clinical Examination and the National Emergency Medicine M4 Exams

Katherine Hiller, MD, MPH*
Joseph House, MD†
Luan Lawson, MD, MAEd‡
Stacey Poznanski, DO§
Thomas K. Morrissey, MD, PhD¶

*University of Arizona School of Medicine, Department of Emergency Medicine, Tucson, Arizona
†University of Michigan School of Medicine, Department of Emergency Medicine, Ann Arbor, Michigan
‡East Carolinas University Brody School of Medicine, Department of Emergency Medicine, Greenville, North Carolina
§Wright State University School of Medicine, Department of Emergency Medicine, Dayton, Ohio
¶University of Florida School of Medicine-Jacksonville, Department of Emergency Medicine, Jacksonville, Florida

Section Editor: Jonathan Fisher, MD
Submission history: Submitted May 1, 2015; Revision received September 7, 2015; Accepted September 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27305

Introduction: In April 2013, the National Board of Medical Examiners (NBME) released an Advanced Clinical Examination (ACE) in emergency medicine (EM). In addition to this new resource, CDEM (Clerkship Directors in EM) provides two online, high-quality, internally validated examinations. National usage statistics are available for all three examinations, however, it is currently unknown how students entering an EM residency perform as compared to the entire national cohort. This information may help educators interpret examination scores of both EM-bound and non-EM-bound students.

Objectives: The objective of this study was to compare EM clerkship examination performance between students who matched into an EM residency in 2014 to students who did not. We made comparisons were made using the EM-ACE and both versions of the National fourth year medical student (M4) EM examinations.

Method: In this retrospective multi-institutional cohort study, the EM-ACE and either Version 1 (V1) or 2 (V2) of the National EM M4 examination was given to students taking a fourth-year EM rotation at five institutions between April 2013 to February 2014. We collected examination performance, including the scaled EM-ACE score, and percent correct on the EM M4 exams, and 2014 NRMP Match status. Student t-tests were performed on the examination averages of students who matched in EM as compared with those who did not.

Results: A total of 606 students from five different institutions took both the EM-ACE and one of the EM M4 exams; 94 (15.5%) students matched in EM in the 2014 Match. The mean score for EM-bound students on the EM-ACE, V1 and V2 of the EM M4 exams were 70.9 (n=47, SD=9.0), 84.4 (n=36, SD=5.2), and 83.3 (n=11, SD=6.9), respectively. Mean scores for non-EM-bound students were 68.0 (n=256, SD=9.7), 82.9 (n=243, SD=6.5), and 74.5 (n=13, SD=5.9). There was a significant difference in mean scores in EM-bound and non-EM-bound student for the EM-ACE ($p=0.05$) and V2 ($p<0.01$) but not V1 ($p=0.18$) of the National EM M4 examination.

Conclusion: Students who successfully matched in EM performed better on all three exams at the end of their EM clerkship. [West J Emerg Med. 2015;16(6):919–922.]

INTRODUCTION

Assessment using a high stakes examination is an important component of a medical student's rotation grade. In the latest State of the Clerkship survey, on average, 25% of a student's grade is determined by a high stakes end-of-rotation examination score.[1] Clerkship directors frequently use the National emergency medicine (EM) fourth year medical student (M4) examination or the National Board of Medical Examiners (NBME) EM Advanced Clinical Examination (ACE) for this assessment.[2-4] These examinations are administered in both required and elective rotations, thus are given to both "career-bound" (i.e. students interested in matching in EM) and "non-career-bound" students.

In addition to providing students and clerkship directors feedback on a student's knowledge base, these examinations provide feedback on how students compare to their peers nationally. Versions 1 (V1) and 2 (V2) of the National EM M4 exams have historic means and standard deviations for examination administrations (www.saemtests.org), while the NBME has reported scaled scores for the EM-ACE since October 2013 (and retrospectively reported them for examination administrations before October 2013).[4] While the examination means and standard deviations vary slightly year to year, the most recent (2014-5) EM M4 V1 mean is 81.5 (SD=3.7) and V2 is 78.4 (SD=4.4). The EM-ACE is scaled to a mean score of 70 (SD=8).[5]

The National EM M4 exams report statistics on the entire population of students who have taken the examination, and the EM-ACE has been scored based on all fourth-year first-time LCME-accredited medical student administrations. Little is known about how students who ultimately match in EM perform on these examinations as compared to their non-EM-bound peers.

The objective of this study was to determine the mean and standard deviation performance of students who matched in EM on the three commonly used exams for student assessment of EM medical knowledge. Additionally, we compared performance of EM-bound and non-EM-bound students on these examinations.

METHODS

We performed this multicenter, retrospective, cohort study as a subset analysis across five U.S. allopathic medical schools between May 2013 and April 2014. During this period, the NBME was attempting to validate the EM-ACE quickly in order to develop scaled scores and the exam was offered free of charge. In order to correlate EM-ACE performance to exams that already had established validity, all fourth-year medical students participating in a fourth-year EM rotation at the study sites were administered both the NBME EM-ACE and one version of an EM M4 examination.[6] The dataset used for this study was derived from the data collected for the EM-ACE National EM M4 correlation project.

The study sites varied with regard to having mandatory selective or elective EM rotations, but were all four weeks in duration and used the standardized curriculum recommended by the Clerkship Directors in EM (CDEM). Study sites administered either V1 or V2 of the EM M4 examination based upon site preference. Exams were taken consecutively, within one day of each other, at the end of the rotation. Individual study sites determined which examination was administered first. Both exams were administered by the same clerkship coordinator or other administrator according to respective protocols developed by the NBME and CDEM. At all sites, students were aware that the EM M4 examination would count towards their grade, as per local institution protocol. Without longitudinal performance data or norms, most sites did not count NBME examination towards the final rotation grade; however, to encourage students to take the NBME examination seriously, some institutions advised students that although the NBME examination could not lower their grade, a strong performance would be reflected in their final evaluation. One institution used the NBME score for a small portion (5%) of the final course grade.

De-identified data were collected by the clerkship director or coordinator, and included institution, NBME scaled score, the version of the EM M4 examination administered (V1, V2) and the score on that examination. After the 2014 National Resident Matching Program (NRMP) Match, whether the student matched in EM (match status) was also collected as a dichotomous variable. Student's t-tests were performed on the examination averages of students who matched in EM as compared with those who did not.

We performed data collection in Microsoft Excel 2007 and data analysis with StataMP 11.0 (College Station, TX).

This project was determined to be exempt from human subjects review by the institutional review boards of all participating institutions.

RESULTS

A total of 606 students took both the EM-ACE and one of the versions of the National EM M4 examination. Of the total cohort, 94 (15.5%) matched into EM in March 2014. This represents 5.3% of all the EM residency positions in the 2014 NRMP Match.[7]

Students who matched in EM had higher examination averages on all three examinations. This difference was statistically significantly for the EM-ACE and Version 2 of the National EM M4 examination ($p=0.05$, $p<0.01$ respectively). See Table.

DISCUSSION

While it is perhaps not surprising that EM-bound students perform better on EM exams than non-EM bound students, this phenomenon has not been described. We report on a small but geographically diverse sample of students who took these exams for the first time. To our knowledge, this is the first time examination means and statistics have been specifically reported for the group of students matching into EM.

Table. Difference in emergecny medicine advanced clinical examination (EM-ACE), EM fourth year medical student (M4) version 1 (V1) and version 2 (V2) examination scores by student match status.

Examination	EM student score (SD)	Non-EM student score (SD)	P value
EM-ACE (scaled score)	70.9 (9.0) N=47	68.0 (9.7) N=256	0.05
V1 EM M4 examination (percent correct)	84.4 (5.2) N=36	82.9 (6.5) N=243	0.18
V2 EM M4 examination (percent correct)	83.3 (6.9) N=11	74.5 (5.9) N=13	<0.01

SD, standard deviation

Such information is valuable to students, advisors and program directors. Students should know how they score in relation to their peers, especially the cohort of EM-bound students, as this information may have a significant impact on their application, interview and match-list strategy. Additionally, clerkship directors and other medical student advisors may be able to use this information to give students an idea of how successfully matched EM-residents performed on their end-of-rotation examination. Finally, this information is valuable to program directors as an objective measure of a candidate's EM knowledge foundation, and may predict future success on other high stakes exams, such as the American Board of EM (ABEM) in-training examination or qualifying certification examination.

LIMITATIONS

Although the study population was taken from five geographically diverse sites, the number of students who matched in EM in this sample was small, a total of 94. This represents 5.3% of all EM spots in the 2014 NRMP match. Match status rather than interest in EM was used to identify the cohorts in part because match status is a discrete and unambiguous variable. The non-EM group consists almost entirely of students who electively pursued specialties other than EM, however, it is likely a small number of students who attempted but were unsuccessful in the EM match are included in this group. We were unable to quantify the number of students in this cohort, as intended specialty match information is only available to the applicant, and may change over time. Additionally, in advising students interested in matching in EM, exam performance compared to successfully matched applicants is a more valuable metric than performance compared to all students attempting to match in EM. Prospective collection of information about intended career goals in relation to examination performance may represent an avenue for future research.

Student scores were likely affected by the perceived importance of the examination. EM-bound students may have prepared more intensely compared to their non-EM peers due to a perceived greater impact on their future career. Sites varied as to whether the clerkship was required, selective or elective. It is possible that non-EM-bound students in an elective/selective rotation might differ from those in a required rotation, in regards to motivation and interest in EM-related material. In addition, site directors used the scores from these exams differently. While students each took the EM-ACE and one of the EM M4 exams, the EM-ACE examination constituted 0-5% of the final rotation grade, and the National EM M4 exams up to 25%. Finally, students in the EM-bound group may have had more experience in EM than their non-EM-bound counterparts prior to their examination.

It is important to note that a knowledge assessment examination is only one measure of student performance. While all these exams are high quality, high stakes, validated exams, they report on only one dimension of a student's capacity to provide EM care. These results must be viewed as one component of the entire application when evaluating a student for residency candidacy.

CONCLUSION

Students who matched into an EM residency performed significantly better on the NBME EM-ACE and Version 2 of the National EM M4 exams. As an objective measure of EM knowledge, these exams may help clerkship directors counsel students about their likelihood of matching into EM. Program directors may be interested in using this information in the evaluation of EM applicants.

Address for Correspondence: Katherine M. Hiller, MD, MPH, 1501 N. Campbell Ave, Tucson, AZ 85724. Email: khiller@aemrc.arizona.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Khandelwal S, Way DP, Wald DA, et al. State of undergraduate education in emergency medicine: a national survey of clerkship directors. *Acad Emerg Med*. 2014;21(1):92-5.
2. Senecal EL, Thomas SH, Beeson MS. A four-year perspective of Society for Academic Emergency Medicine tests: an online testing tool for medical students. *Acad Emerg Med*. 2009;16Suppl2:S42-45.
3. Senecal EL, Heitz C, Beeson MS. Creation and implementation of a national emergency medicine fourth-year student examination. *J*

Emerg Med. 2013;45(6):924-34.

4. NBME. 2014-2015 Subject Examination Fees. 2014. Accessed Jun 7, 2014.
5. Miller ES. National EM M4 exam averages2105.
6. Hiller K, Miller ES, Lawson L, et al. Correlation of the NBME advanced clinical examination in EM and the national EM M4 exams. *West J Emerg Med*. 2015;16(1):138-42.
7. NRMP. Results and Data 2014 Main Residency Match. 2014; Available at: http://www.nrmp.org/wp-content/uploads/2014/04/Main-Match-Results-and-Data-2014.pdf. Accessed May 1, 2015.

Competency Assessment in Senior Emergency Medicine Residents for Core Ultrasound Skills

Jessica N. Schmidt, MD, MPH*
John Kendall, MD†
Courtney Smalley, MD‡

*University of Wisconsin School of Medicine and Public Health, Department of Emergency Medicine, Madison, Wisconsin
†Denver Health Medical Center, Department of Emergency Medicine, Denver, Colorado
‡University of Colorado School of Medicine, Department of Emergency Medicine, Denver, Colorado

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted September 4, 2015; Accepted September 16, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.28587

Introduction*:* Quality resident education in point-of-care ultrasound (POC US) is becoming increasingly important in emergency medicine (EM); however, the best methods to evaluate competency in graduating residents has not been established. We sought to design and implement a rigorous assessment of image acquisition and interpretation in POC US in a cohort of graduating residents at our institution.

Methods: We evaluated nine senior residents in both image acquisition and image interpretation for five core US skills (focused assessment with sonography for trauma (FAST), aorta, echocardiogram (ECHO), pelvic, central line placement). Image acquisition, using an observed clinical skills exam (OSCE) directed assessment with a standardized patient model. Image interpretation was measured with a multiple-choice exam including normal and pathologic images.

Results*:* Residents performed well on image acquisition for core skills with an average score of 85.7% for core skills and 74% including advanced skills (ovaries, advanced ECHO, advanced aorta). Residents scored well but slightly lower on image interpretation with an average score of 76%.

Conclusion: Senior residents performed well on core POC US skills as evaluated with a rigorous assessment tool. This tool may be developed further for other EM programs to use for graduating resident evaluation. [West J Emerg Med. 2015;16(6):923–926.]

INTRODUCTION

Quality resident education in point-of-care ultrasound (POC US) has become increasingly important in the practice of emergency medicine (EM). Several guidelines have been proposed for US curriculum and training.[1,2] Most recently, in 2008 the Council of EM Residency Directors (CORD) together with leaders in the field of POC US published guidelines for minimum education standards. This consensus group made recommendations regarding methods for competency assessment, including a practical examination with direct assessment of US skills and an assessment of image interpretation.[3]

Previous research has shown that there are reliable methods to assess US skills in trainees including using observed clinical skills exam (OSCE),[4-8] written image interpretation,[9,10] and protected hands-on training.[11] However, despite validation of these methods, it remains unclear how best to evaluate residents in POC US. In 2013, Amini et al published a survey study reporting wide variation in current EM residency practices in competency assessment demonstrating that 21% used standardized direct observation tools (SDOTS), a third used multiple choice questions, and another third administered practical exams.[12]

Despite the development of minimum education standards

for POC US, it is unclear whether residents are graduating with the required skill set. We sought to create a rigorous assessment tool using previously validated methods including an OSCE and written exam to evaluate both image acquisition and image interpretation in residents nearing graduation. The aim of our study was to assess how such an assessment tool could aid in evaluating senior residents in five core US skills as defined by the 2008 CORD document. This is, to our knowledge, the first study since publication of the CORD recommendations to describe a methodology for senior resident assessment in POC US.

METHODS

We conducted the study at an urban academic emergency department with an annual census of 55,000. Nine senior residents in a four-year accreditation council for graduate medical education accredited EM residency program participated in the study. Participation was voluntary and the study was approved by the institutional review board with written consent obtained from participants.

All residents completed a two-week US rotation during their first year, acquired a minimum of 150 scans and participated in didactics throughout residency, in addition to using POC US during their clinical experience. One resident also completed a two-week elective in US. On average, residents had spent 37 months in residency at the time of evaluation.

We evaluated residents in US image acquisition and image interpretation for five core EM US applications including echocardiogram (ECHO), aorta, focused assessment with sonography for trauma (FAST), pelvic (trans-abdominal and trans-vaginal) and central line placement.

Image Acquisition

To assess image acquisition, residents were asked to perform five basic POC US skills on a live standardized patient model while two independent US-trained EM physicians completed a pre-defined OSCE checklist regarding their performance. Checklists for ECHO, aorta, FAST and pelvic exams were created by the Academy of Emergency US and are published in the list of CORD Assessment Methods.[13] We obtained the checklist for central line placement from a previously published checklist.[14] Table 1 shows the itemized components of each checklist scored as one point each. As each resident was scored by two examiners, there was a total possible score of 126 points. Results were coded so that the authors of this paper were blinded to the individual reviewer and the resident participant. Residents were given 45 minutes to complete all exams.

Image Interpretation

To assess image interpretation, residents were asked to complete a computer-based multiple-choice quiz. This quiz contained live cine-clips and still images with normal and abnormal pathology. The question bank was created by www.emsono.com. Table 2 outlines the concepts tested and the

Table 1. Point-of-care ultrasound exam type and area of evaluation scored as one point each for objective structured clinical examination.

Exam type	Area of evaluation
ECHO (15 points)	Correct transducer selection
	Identifies parasternal long axis view with RV, LA, LV
	Measures aortic outflow tract
	Identifies parasternal short axis view with RV, LV
	Identifies apical four-chamber view with RV, LV, RA, LA
Aorta (18 points)	Correct transducer selection
	Identifies target and inferior vena cava by compression and/or doppler
	Obtains trans. proximal, middle, distal and bifurcation views
	Obtains longitudinal aorta view
	Performs correct measurement of aorta
	Identifies vertebral body
FAST (11 points)	Correct transducer selection
	Subxiphoid view
	Identifies right upper quadrant with Morisons, tip of liver, inf. pole kidney
	Identifies left upper quadrant view
	Identifies the splenorenal recess
	Visualizes the inferior pole of the left kidney
	Identifies pelvic view in transverse and sagittal planes
Pelvic (12 points)	Correct transducer selection
Trans-abdominal	Obtains uterus view in long and short axis
	Scans through bilateral ovaries in two planes
	Explains how to calculate the fetal heart rate with M-mode
Trans-vaginal	Obtains coronal and sagittal uterus view
Central line (7 points)	Correct transducer selection
	Explains probe positioning/marker orientation
	Identifies target and associated artery
	Measures depth of vein
	Appropriately demonstrates needle entry/ angle of insertion

ECHO, echocardiogram; *RV*, right ventricle; *LA*, left atrium; *LV*, left ventricle; *RA*, right atrium; *FAST*, focused assessment with sonography for trauma

breakdown of normal and abnormal pathology. Results of the quiz were compiled by an external server and presented in a de-identified data set to the study authors.

RESULTS

Image Acquisition

For the image acquisition, the average total score from the OSCE checklist was 92 out of 126 (74%) with a range from 75 to 110 (61-87%). When excluding advanced US competencies including evaluation of ovaries, advanced ECHO, and aortic root measurement, the average score was 85.7 out of 100 (85.7%).

Central line/FAST*:* Residents scored highest in the central line application with all residents scoring 100% correct. They also scored highly in the FAST exam with an average of score of 21.3 out of 22 (97%)

Aorta: On assessment of the abdominal aorta, residents scored an average of 17.7 out of 20 points (88%). On advanced aortic imaging, however, only three residents correctly identified the superior mesenteric artery and/or the celiac trunk and one resident identified the spinal stripe.

Pelvic: For the trans-abdominal and trans-vaginal pelvic assessments, the average scores were 10 out of 12 (85%) and 9.6 out of 12.0 (80%) respectively. The most frequently missed structures on trans-abdominal US were the left and right ovaries (scored one point each). For the trans-vaginal US, in addition to missed ovaries, two of the residents reversed the coronal and sagittal orientations.

ECHO: For ECHO evaluation, resident averaged a score of 18.7 out of 30 (62%). Two residents had outlying low scores of 4 and 9 due to inability to obtain an apical four chamber and parasternal short axis views. Six residents incorrectly identified or incorrectly measured the aortic root and four incorrectly identified the chambers on parasternal long-axis view.

Image Interpretation

For image interpretation, the average score was 76% with a range from 68 to 89%. One resident data point was excluded due to an incomplete on-line quiz due to technical errors. The majority of the scores ranged from 71-79 percent and only one score was below 70%.

DISCUSSION

Our study demonstrates that senior residents performed well at image acquisition in several applications including central line placement, FAST, and basic aorta. There was more variation in our cohort in pelvic and advanced ECHO and aorta. For pelvic image acquisition, residents primarily had difficulty identifying the ovaries, and for ECHO there was variation in ability to obtain parasternal-short and apical four chamber views. For aorta, there was difficulty with the superior mesenteric artery and spinal stripe. Adnexal pathology, heart chamber size and comparison, and aortic anatomy are all considered "advanced skills" by CORD. Because the OSCE included both "core" and "advanced" skills as defined by CORD, the limitations noted by our resident cohort may not necessarily mean they are not meeting "core skills," but may reflect deficits in more advanced POC US skills. When the data was re-analyzed excluding those aspects of the exams that were considered advanced skills, overall scores went from 74 to 85.7%. For future studies, researchers may want to consider an edited version of the CORD OSCE that includes only core skills.

Table 2. Question categories for image-interpretation quiz.

Category	Subcategory	Questions (# abnormal)
AAA	N/A	5 (2)
FAST	N/A	5 (2)
ECHO	Pericardial effusion	5 (2)
	RV: LV	5 (2)
	Ejection Fraction	5 (2)
Pelvic	Pregnancy	6 (3)
	Yolk sac/gestational sac/fetal pole	2
	Positive FAST with no IUP and +UPT	2
Central line	Identifies needle tip vs artifact Identifies correct vessel	3
Total		38

AAA, abnormal aortic aneurysm; *FAST*, focused assessment with sonography for trauma; *ECHO*, echocardiogram; *RV*, right ventricle; *LV*, left ventricle; *IUP*, intrauterine pregnancy; *UPT*, urinary pregnancy test

Overall, there was a tendency for residents to score higher on the image acquisition than image interpretation. This may be due to the method of scoring for the two different testing modalities. Additionally, residents have been exposed to hands-on US during their clinical training, primarily identifying normal structures. It may be that more exposure to pathology, as tested in the image interpretation section, is required.

Although our residents performed well on testing overall, a score that reflects "competency" has yet to be defined. Leaders in the field of POC US have suggested that a comprehensive approach is needed to fully assess an individual resident's competency in POC US.[3,15] Although our study offers a rigorous tool using validated methods, further studies are warranted to evaluate how performance on these measures correlates to clinical performance.

LIMITATIONS

One potential limitation was possible reviewer bias. Due to the nature of the OSCE, the two reviewers were not blinded to the residents and may have had previous clinical experience with them, which may have skewed their interpretation of

image acquisition. To best compensate for this limitation, two reviewers were selected instead of one to ensure that there was consistency. Future studies may use independent US-trained reviewers to limit this bias.

Additionally, this study was performed on a small cohort of residents at a single institution. While we were able to identify areas of weakness overall for our program and for individual residents, it remains to be seen how a similar assessment would work at other institutions.

Finally, with the OSCE exams put forth by CORD, there is no recommendation for how many items on the checklist indicate "competency" or how these scores correlate with clinical performance. Reproducing this study across institutions may allow educators to define an acceptable score for competency.

CONCLUSION

This is the first paper to measure POC US skills in senior residents using a rigorous methodology to assess both image acquisition and image interpretation in core EM applications as defined by CORD. Overall, we found that senior residents performed well on image interpretation but had difficulty with image acquisition in more advanced US applications. Further work by other institutions and leaders in US is needed to translate performance on these core measures to overall clinical performance.

Address for Correspondence: Jessica N. Schmidt, MD, MPH, Berbee Walsh Department of Emergency Medicine, 800 University Bay Drive, Suite 310, Mail code 9123, Madison, WI 53705. Email: jessica.schmidt@denverem.org.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Mateer J, Plummer D, Heller M. Model curriculum for physician training in emergency ultrasonography. *Ann Emerg Med.* 1994;23:95-102.
2. ACEP emergency ultrasound guidelines-2001. *Ann of Emerg Med.* 2001;38:470-81.
3. Akhtar S, Theodoro D, Gaspari R, et al. Resident Training in Emergency Ultrasound: Consensus Recommendations from the 2008 Council of Emergency Medicine Residency Directors Conference. *Acad Emerg Med.* 2009;16:S32-36.
4. Sisley A, Johnson S, Erickson W, et al. Use of an Objective Structured Clinical Examination (OSCE) for the assessment of physician performance in the ultrasound evaluation of trauma. *J Trauma*. 1999;47:627-31.
5. Kissin E, Grayson P, Cannella A, et al. Musculoskeletal ultrasound objective structured clinical examination: an assessment of the test. *Arthritis Care Res (Hoboken)*. 2014;66:2-6.
6. Breitkreutz R, Dutiné M, Scheiermann P, et al. Thorax, trachea, and lung ultrasonography in emergency and critical care medicine: assessment of an objective structured training concept. *Emerg Med Int*. 2013:312758.
7. Hofer M, Kamper L, Sadlo M, et al. Evaluation of an OSCE assessment tool for abdominal ultrasound courses. *Ultraschall Med*. 2011;32:184-90.
8. Knobe M, Münker R, Sellei RM, et al. Peer teaching: a randomised controlled trial using student-teachers to teach musculoskeletal ultrasound. *Med Educ.* 2010;44:148-55.
9. Mandavia D, Aragona J, Chan L, et al. Ultrasound training for emergency physicians--a prospective study. *Acad Emerg Med.* 2000;7:1008-14.
10. Blackstock U, Munson J, Szyld D. Bedside ultrasound curriculum for medical students: report of a blended learning curriculum implementation and validation. *J Clin Ultrasound.* 2015;43:139-44.
11. Noble V, Nelson B, Sutingco A, et al. Assessment of knowledge retention and the value of proctored ultrasound exams after the introduction of an emergency ultrasound curriculum. *BMC Med Educ*. 2007;7.
12. Amini R, Adhikari S, Fiorello A. Ultrasound Competency Assessment in Emergency Medicine Residency Programs. *Acad Emerg Med.* 2014;21:799-801.
13. Ultrasound PC12 milestone workgroup. Academy of Emergency Ultrasound. Available at: http://emmilestones.pbworks.com/w/page/66439892/Ultrasound%20PC12%20milestone%20workgroup. Accessed Jun 12, 2013.
14. Rosen BT, Uddin PQ, Harrington AR, et al. Does personalized vascular access training on a nonhuman tissue model allow for learning and retention of central line placement skills? Phase II of the procedural patient safety initiative (PPSI-II). *J Hosp Med*. 2009;4:423–9.
15. Diller D and Bramante R. Ask the Expert: How do you evaluate trainee competency in Point of Care Ultrasound? An interview with Resa Lewiss. Available at: http://www.acep.org/Content.aspx?ID=98246. Accessed Aug 25, 2015.

Mentoring during Medical School and Match Outcome among Emergency Medicine Residents

Erin Dehon, PhD
Margaret H. Cruse, MS
Brandon Dawson, MD
Loretta Jackson-Williams, MD

University of Mississippi Medical Center, Department of Emergency Medicine, Jackson, Mississippi

Section Editor: Jonathan Fisher, MD
Submission history: Submitted April 24, 2015; Revision received October 5, 2015; Accepted September 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27010

Introduction: Few studies have documented the value of mentoring for medical students, and research has been limited to more subjective (e.g., job satisfaction, perceived career preparation) rather than objective outcomes. This study examined whether having a mentor is associated with match outcome (where a student matched based on their rank order list [ROL]).

Methods: We sent a survey link to all emergency medicine (EM) program coordinators to distribute to their residents. EM residents were surveyed about whether they had a mentor during medical school. Match outcome was assessed by asking residents where they matched on their ROL (e.g., first choice, fifth choice). They were also asked about rank in medical school, type of degree (MD vs. DO), and performance on standardized tests. Residents who indicated having a mentor completed the Mentorship Effectiveness Scale (MES), which evaluates behavioral characteristics of the mentor and yields a total score. We assessed correlations among these variables using Pearson's correlation coefficient. Post-hoc analysis using independent sample t-test was conducted to compare differences in the MES score between those who matched to their first or second choice vs. third or higher choice.

Results: Participants were a convenience sample of 297 EM residents. Of those, 199 (67%) reported having a mentor during medical school. Contrary to our hypothesis, there was no significant correlation between having a mentor and match outcome ($r=0.06$, $p=0.29$). Match outcome was associated with class rank ($r=0.13$, $p=0.03$), satisfaction with match outcome ($r= -0.37$, $p<0.001$), and type of degree ($r=0.12$, $p=0.04$). Among those with mentors, a t-test revealed that the MES score was significantly higher among those who matched to their first or second choice ($M=51.31$, $SD=10.13$) compared to those who matched to their third or higher choice ($M=43.59$, $SD=17.12$), $t(194)=3.65$, $p<0.001$, $d=0.55$.

Conclusion: Simply having a mentor during medical school does not impact match outcome, but having an effective mentor is associated with a more favorable match outcome among medical students applying to EM programs. [West J Emerg Med. 2015;16(6):927–930.]

INTRODUCTION

Mentoring has been associated with numerous benefits for individuals working in fields ranging from business to academic medicine. Among academic physicians, mentoring is associated with increased job satisfaction, higher salary, increased research productivity, and career advancement.[1-3] Physicians with mentors were found to be 2.3 times more likely to be promoted than those without mentors.[4]

Mentoring has demonstrated similar benefits for medical trainees including both residents and medical students. Compared to non-mentored residents, mentored residents were twice as likely to state that they received excellent career preparation.[5] A systematic review of mentoring programs for medical students revealed that having a mentor is associated with increased research productivity and interest in academic careers, enhanced well-being, and specialty choice for medical students.[6]

Despite the aforementioned benefits of mentoring for medical trainees, few studies, with the exception of those focused on research productivity, have examined quantifiable (vs. subjective) benefits of mentoring for medical trainees. Furthermore, few studies have examined the value of mentoring among medical students who enter emergency medicine (EM). The purpose of this study was to examine EM residents' experience of mentorship during medical school and its relationship to match outcome. Our hypothesis is that EM residents who report having a mentor during medical school will be more likely to have matched to a residency program at the top of their rank order list (ROL).

METHODS

Participants were recruited through the EM Association of Residency Coordinators listserv. An email with information about the study and a link to the survey was sent to the program coordinators who were asked to distribute the email to their EM residents. The institutional review board approved this study and a waiver of signed consent.

EM residents were surveyed about whether they had a mentor during medical school using the definition provided by Ramanan[5]: "an active partner in an ongoing relationship who helps you maximize your potential and achieve your personal and professional goals." Residents also reported their rank in medical school, degree (MD vs. DO), location of medical school (U.S. or international) and performance on standardized tests. For the purpose of this study, we assessed match outcome by asking residents where they matched based on their ROL (e.g., first choice, fifth choice). Resident satisfaction with match outcome was measured using a five-point scale (very dissatisfied to very satisfied).

Residents who indicated having a mentor were directed to complete the Mentorship Effectiveness Scale (MES).[7] The MES is a 12-item self-report measure designed to assess the overall effectiveness of mentoring. Each item describes behavioral characteristics of a mentor, which are rated using a five-point Likert-type scale (0=strongly disagree, 5=strongly agree) or "NA" if the item did not apply. Lastly, residents reported length of relationship with mentor, gender of mentor, and whether or not they still communicated with their mentor.

We used chi-square analyses to compare applicant characteristics (e.g., sex, United States Medical Licensing Examination [USMLE] score, rank in medical school) of those with and without mentors. Pearson correlations were conducted to examine the relationship between having a mentor and match outcome. We conducted post-hoc analysis using an independent sample t-test to compare differences in the MES score between those who matched to one of their top two choices vs. third or higher choice.

RESULTS

The convenience sample was 297 EM residents; 59% (n=176) were male and 41% (n=121) were female. The majority were allopathic (79%, n=235) and U.S. graduates (93%, n=277). These characteristics are largely consistent with the National Resident Matching Program data.[8] About two-thirds (67%) reported having a mentor during medical school. Males (66%, n=117) and females (67%, n=82) reported having a mentor during medical school. Of those with mentors, 76% (n=148) reported that their mentor was self-identified versus assigned by their school 24% (n=46). Most mentors were EM physicians (80%, n=159). Male mentors (72%, n=140) were more common than female mentors (28%, n=55). About half (55%, n=110) reported that they still communicated with their mentor.

A comparison of those with and without mentors is presented in Table 1. There was a significant association between type of degree and mentorship, $\chi^2(1, n=297)=6.73$, $p<0.01$ with the odds of having a mentor 2.1 times higher among those with an allopathic degree. There was also a significant association between location of medical school and mentorship, χ^2 $(1, n=297)=6.73$, $p<0.05$. The odds of having a mentor were 3.4 times higher for those who attended a U.S. school.

Regarding match outcome, the majority of respondents reported matching to their first choice (n=176, 59%), followed by second (n=56, 19%), third (n=27, 9%), fourth (5%, n=15), fifth (2%, n=7), sixth (1%, n=4), seventh (1%, n=3), and 8th or higher choice (3%, n=8). Contrary to our hypothesis, there was no significant correlation between having a mentor and match outcome ($r=0.06$, $p=0.29$). A nearly equal number of respondents with and without mentors matched to one of their top two choices (Table 1). Match outcome was significantly associated with class rank ($r=0.13$, $p=0.03$), satisfaction with match outcome ($r= -0.37$, $p<0.001$), and having an MD (vs. DO) ($r=0.12$, $p=0.04$). USMLE was not significantly associated with match outcome.

Among those with mentors, we used a t-test to compare MES scores among EM residents who matched to one of their top two choices to those who matched lower on their list. The MES score was significantly higher among those who matched to their first or second choice ($M=51.31$, $SD=10.13$) compared to those who matched lower on their ROL ($M=43.59$, $SD=17.12$), $t(194)=3.65$, $p<0.001$, $d=0.55$. Table 2 shows the descriptive statistics for each of the items on the MES. Residents who still communicated with their mentor were more likely to have matched to their first or second choice, χ^2 $(1, n=198)=10.79$, $p<0.01$.

Table 1. Characteristics of residents with and without mentors during medical school.

Characteristic	Mentor (n=199)	No mentor (n=98)	Chi-square	P
Sex [n (%)]			0.05	0.82
Male	117 (58.8%)	59 (60.2%)		
Female	82 (41.2%)	39 (39.8%)		
USMLE score [n (%)]			3.27	0.35
181-210	18 (9.3%)	13 (13.7%)		
211-250	131 (67.5%)	59 (62.1%)		
>250	36 (18.6%)	15 (15.8%)		
Did not take USMLE	9 (4.6%)	8 (8.4%)		
Rank in medical school [n (%)]			2.50	0.65
Top sixth	42 (21.1%)	26 (27.1%)		
Top third	48 (24.1%)	25 (26.0%)		
Middle third	59 (29.6%)	27 (28.1%)		
Bottom third	11 (5.5%)	3 (3.1%)		
Not used by medical school	39 (19.6%)	15 (15.6%)		
Degree [n (%)]			6.73	0.009**
MD (allopathic)	166 (83.4%)	69 (70.4%)		
DO (osteopathic)	33 (16.6%)	29 (29.6%)		
Medical school location [n (%)]			6.72	0.01*
United States	190 (96.4%)	87 (88.8%)		
International	7 (3.6%)	11 (11.2%)		
Match outcome [n (%)]			0.71	0.39
1st or 2nd choice	158 (79.8%)	74 (75.5%)		
3rd choice or higher	40 (20.2%)	24 (24.5%)		
Satisfaction with match [n (%)]			8.56	0.07
Very dissatisfied	6 (3.1%)	4 (4.2%)		
Dissatisfied	4 (2.1%)	1 (1.0%)		
Neutral	9 (4.6%)	4 (4.2%)		
Satisfied	33 (16.9%)	30 (31.3%)		
Very satisfied	143 (73.3%)	57 (59.4%)		

USMLE, United States medical licensing exam
*Significant at p<0.05.
**Significant at p<0.01.

DISCUSSION

Findings do not provide support for the hypothesized relationships between having a mentor during medical school and matching high on one's rank list. Match outcome was more likely to be associated with other factors including class rank and type of degree (i.e., MD). Nonetheless, we did find a relationship between degree of mentorship effectiveness and match outcome. Specifically, we found that among EM residents with mentors, those who reported greater mentor effectiveness were more likely to match to their first or second choice. Taken together, these findings suggest having a supportive, motivating, and helpful mentor may be one of many factors that can increase an applicant's chance of matching to one of their top choices.

This study also sheds light on the prevalence of mentoring among medical school students who enter EM. Two-thirds of the respondents reported having a mentor during medical school. This number is much higher than previous reports, which found the prevalence of mentorship among medical students to be 50%.[2] It is unclear whether this finding reflects a genuine increase in mentoring, a higher prevalence of mentoring among students interested in EM, or some other anomaly. Given that the definitions of mentoring often vary from study to study, it is often difficult to make meaningful comparisons across studies. We also found that graduates of osteopathic and international graduates were less likely to have mentors compared to allopathic and U.S. graduates. It may be that mentoring programs are more prevalent in U.S.

Table 2. Descriptive statistics of mentor effectiveness and match outcome. Means for the Mentoring Effectiveness Scale items are based on a five-point scale from 0 (strongly disagree) to 5 (strongly agree). NA was an option and was coded as 0.

Variable	Mean	SD
My mentor was accessible.	4.38	0.95
My mentor demonstrated professional integrity.	4.66	0.64
My mentor demonstrated content expertise in my area of need.	4.47	0.90
My mentor was approachable and easy to talk with about concerns.	4.58	0.84
My mentor was supportive and encouraging.	4.58	0.86
My mentor provided constructive and useful critique of my work.	4.24	1.01
My mentor motivated me to improve my work product.	4.25	1.03
My mentor was helpful in providing direction and guidance on professional issues (e.g., networking).	4.00	1.26
My mentor answered my questions satisfactorily (e.g., timely response, clear, comprehensive).	4.47	0.93
My mentor was helpful in providing advice on work/school and personal life.	4.04	1.30
My mentor suggested appropriate resources (e.g., experts, contacts, source materials).	4.22	1.06
My mentor challenged me to extend my abilities (e.g., risk taking, try a new activity, draft a section of an article).	3.98	1.26
Total score	49.79	12.17

SD, standard deviation

allopathic schools, although most respondents (76%) reported that their mentor was self-identified, not assigned by the medical school.

LIMITATIONS AND FUTURE DIRECTIONS

First, a major concern is the relatively small sample size and use of a convenience sample. Furthermore, given the design of our study we were unable to include unmatched applicants which could have had a significant impact on the results. Second, these results are based on unverified self-report data. Future studies would benefit from the inclusion of other methods of data collection, such as a review of medical school records for verification of data. Third, the finding that most mentors were self-identified raises the question of whether students who seek out mentors have personal characteristics (e.g., motivation) that contribute to their success. Fourth, given the correlational nature of the study, it is impossible to determine the exact nature of the relationship between mentorship and match outcome. Future studies that incorporate pre and post designs and/or random assignment of students to mentors are needed to more fully examine the relationship between mentorship and match outcome. Lastly, although we found a significant difference between the MES scores for higher vs. lower matching students, more research is needed to verify the meaningfulness of these results.

CONCLUSION

These results suggest that simply having a mentor during medical school does not impact match outcome but having an effective mentor is associated with a more favorable match outcome among medical students applying to EM programs.

Address for Correspondence: Erin Dehon, PhD, University of Mississippi Medical Center, 2500 N State St., Jackson, MS 39216. Email: edehon@umc.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Jackson VA, Palepu A, Szalacha L, et al. "Having the right chemistry": a qualitative study of mentoring in academic medicine. *Acad Med*. 2003;78(3):328-34.
2. Sambunjak D, Straus SE, Marusic A. Mentoring in academic medicine: a systematic review. *JAMA*. 2006;296(9):1103-15.
3. Garmel GM. Mentoring medical students in academic emergency medicine. *Acad Emerg Med*. 2004;11(12):1351-7.
4. Wise MR, Shapiro H, Bodley J, et al. Factors affecting academic promotion in obstetrics and gynaecology in Canada. *J Obstet Gynaecol Can*. 2004;26(2):127-36.
5. Ramanan RA, Taylor WC, Davis RB, et al. Mentoring matters. Mentoring and career preparation in internal medicine residency training. *J Gen Intern Med*. 2006;21(4):340-5.
6. Frei E, Stamm M, Buddeberg-fischer B. Mentoring programs for medical students--a review of the PubMed literature 2000-2008. *BMC Med Educ*. 2010;10:32.
7. Berk RA, Berg J, Mortimer R, et al. Measuring the effectiveness of faculty mentoring relationships. *Acad Med*. 2005;80(1):66-71.
8. The Match. Results and Data: 2014 Main Resident Match. Washington, DC: National Resident Matching Program; 2014: 1,8,15,35-39.

BREIF RESEARCH REPORT

Emergency Medicine Residents Consistently Rate Themselves Higher than Attending Assessments on ACGME Milestones

Katja Goldflam, MD
Jessica Bod, MD
David Della-Giustina, MD
Alina Tsyrulnik, MD

Yale University School of Medicine, Department of Emergency Medicine, New Haven, Connecticut

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted April 29, 2015; Revision received August 11, 2015; Accepted August 25, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.27247

Introduction: In 2012 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which implemented milestones to assess the competency of residents and fellows. While attending evaluation and feedback is crucial for resident development, perhaps equally important is a resident's self-assessment. If a resident does not accurately self-assess, clinical and professional progress may be compromised. The objective of our study was to compare emergency medicine (EM) resident milestone evaluation by EM faculty with the same resident's self-assessment.

Methods: This is an observational, cross-sectional study that was performed at an academic, four-year EM residency program. Twenty-five randomly chosen residents completed milestone self-assessment using eight ACGME sub-competencies deemed by residency leadership as representative of core EM principles. These residents were also evaluated by 20 faculty members. The milestone levels were evaluated on a nine-point scale. We calculated the average difference between resident self-ratings and faculty ratings, and used sample t-tests to determine statistical significance of the difference in scores.

Results: Eighteen residents evaluated themselves. Each resident was assessed by an average of 16 attendings (min=10, max=20). Residents gave themselves statistically significant higher milestone ratings than attendings did for each sub-competency examined ($p<0.0001$).

Conclusion: Residents over-estimated their abilities in every sub-competency assessed. This underscores the importance of feedback and assessment transparency. More attention needs to be paid to methods by which residency leadership can make residents' self-perception of their clinical ability more congruent with that of their teachers and evaluators. The major limitation of our study is small sample size of both residents and attendings. [West J Emerg Med. 2015;16(6):931–935.]

INTRODUCTION

In 2012 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which implemented milestones to assess the abilities and progress of residents. Each milestone is a significant, progressive, competency-based point in the development of a resident. These milestones evaluate accomplishments that identify specialty-specific knowledge, skills, attitudes and behaviors that can be used as outcome measures within the general competencies.[1,2] Emergency medicine (EM) has developed 23 sub-competencies, with five milestone levels within each. Residents are expected to progress through levels of proficiency as they complete their training.[2,3]

Attending evaluation and feedback is crucial for resident development. However, at least as equally important is

a resident's self-assessment. This is because feedback from others is often interpreted and integrated through the framework of a learner's self-assessment.[4,5] Learners use an amalgam of self-assessment and feedback to generate actionable goals for improvement.[6] If a resident does not accurately self-assess, clinical and professional progress may be compromised. A resident who is unable to accurately judge his or her own abilities may fail to achieve the necessary skills to be a safe and effective physician. In other words, failure to acknowledge deficiencies may lead to a failure to correct them.

Attending physicians working at academic centers are used as the benchmark in assessing a resident's abilities as a physician. To date, no study has compared resident self-assessment to attending assessment using the standardized framework of the ACGME milestones.

Goals of this Investigation

Our study used the framework of the ACGME milestones to compare EM resident evaluation by EM faculty with the same residents' self-assessments.

METHODS

This study is an observational, cross-sectional study performed at an academic EM residency. A human investigation committee (HIC) exemption was granted by the institutional IRB. All residents from EM post graduate year (PGY) 1 through 4 level were included in the study, with the exception of the single resident who helped to conduct the study. Twenty-five residents were chosen using a random number generator to participate in the study. The remaining residents were omitted due to time limitations on attendings filling out the forms and concerns that too large a number of evaluations would be prohibitive to attending willingness to participate in the study.

These residents completed self-assessments of milestone levels using eight ACGME sub-competencies that were chosen as representative of core EM principles by residency leadership consensus. Moreover, residency leadership agreed that a large group of attending evaluators would likely be able to comment on these, more familiar, sub-competencies for the majority of residents. The residency leadership consensus consisted of the residency program director and associate program directors. These included Emergency Stabilization (PC1), History and Physical (PC2), Diagnostic Studies (PC3), Diagnosis (PC4), Disposition (PC7), Communication (ICS1), Multi-Tasking (PC8), and Team Management (ICS2). These same residents were also evaluated by 20 faculty members using identical milestones. Faculty members have contact with residents in various settings, which include clinical shifts, simulation laboratory, and in small-group teaching sessions. Faculty members were able to opt out of assessing any resident whom they felt they could not evaluate due to limited interaction. The sub-competencies were evaluated on a nine-point scale, which reflects the rubric published by the ACGME (Figure). No advanced training or instruction was provided regarding the utilization of the ACGME milestones. No other evaluation tools were provided to faculty when they were asked to assign a score.

We calculated the average difference between resident self-ratings and faculty ratings. Sample t-tests were used to determine the statistical significance of the difference in scores. We carried out mixed models analyses to determine if there were any significant interactions between the rater type (self vs. attending) and program year. For each program year, we calculated and compared the difference in the least square means between residents and their attending raters to the overall difference in least square means for each sub-competency.

RESULTS

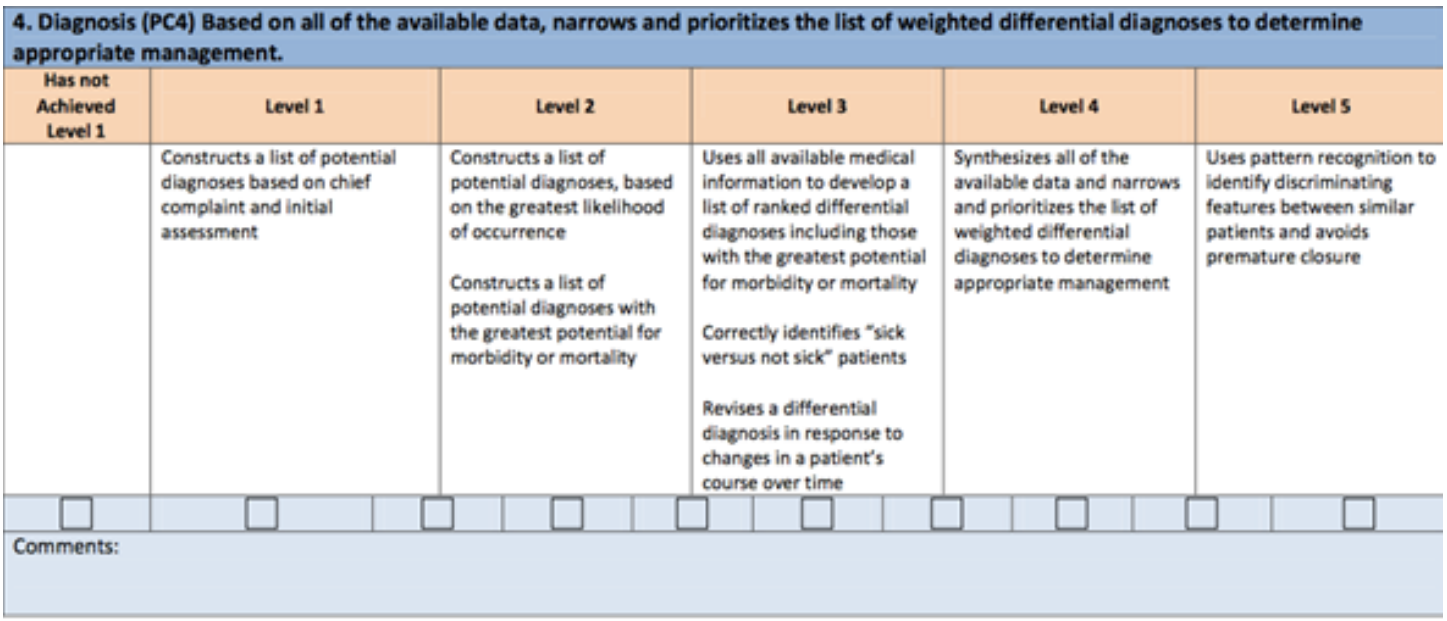
Eighteen of the 25 residents surveyed completed the evaluation. Each resident was assessed by an average of 16 attendings (min=10, max=20). Residents gave themselves higher milestone ratings than attendings did for each of the eight sub-competencies evaluated (Table 1). The mean difference in score for each sub-competency was close to one point, with the exception of "Team Management," which was 0.5 points. For seven out of eight sub-competencies, the difference in resident milestone self-assessment score and attending milestone assessment score was statistically significant ($p<0.05$). The one sub-competency where statistical significance was not reached was "Team Management" (p=0.09).

Mixed model analysis showed statistically significant differences between self-ratings and attending ratings in most sub-competencies for the PGY 1 and 3 cohorts (Table 2 and Table 3). The PGY 2 cohort had fewer differences across sub-competencies, with statistically significant differences in only three sub-competencies (Table 4). For PGY 4, self and attending ratings did not significantly differ in any sub-competency (Table 5).

DISCUSSION

Our study found that residents (combined PGY1 through PGY4) consistently rated themselves as more proficient for each sub-competency than did their attending evaluators. This is consistent with prior data showing that physician self-assessment typically does not correlate with external measures of performance.[7] Although self-assessment may be inaccurate, it is important for evaluators to consider learner self-image when giving feedback. This feedback will undoubtedly be interpreted by the learner through a filter of his/her own perception.[8] For example, feedback from an attending that is lower than a learner feels he/she attained, may be rejected by that learner who believes he or she has reached a higher level of proficiency. This could negatively impact the development and growth of that learner.

Our study illustrates that milestone-based assessment



4. Diagnosis (PC4) Based on all of the available data, narrows and prioritizes the list of weighted differential diagnoses to determine appropriate management.					
Has not Achieved Level 1	Level 1	Level 2	Level 3	Level 4	Level 5
	Constructs a list of potential diagnoses based on chief complaint and initial assessment	Constructs a list of potential diagnoses, based on the greatest likelihood of occurrence Constructs a list of potential diagnoses with the greatest potential for morbidity or mortality	Uses all available medical information to develop a list of ranked differential diagnoses including those with the greatest potential for morbidity or mortality Correctly identifies "sick versus not sick" patients Revises a differential diagnosis in response to changes in a patient's course over time	Synthesizes all of the available data and narrows and prioritizes the list of weighted differential diagnoses to determine appropriate management	Uses pattern recognition to identify discriminating features between similar patients and avoids premature closure
Comments:					

Figure. Sample Accreditation Council for Graduate Medical Education milestone used to assess competency of emergency medicine residents and fellows.

Table 1. Comparison of all residents' post graduate years 1-4 self-rating to attending rating.

Evaluation construct	Self	Rater(s)	Estimated difference ± standard error	95% CL	*p*
Communication	6.68 ± 0.33	5.54 ± 0.10	1.14 ± 0.32	(0.49, 1.78)	0.0006
Diagnosis	6.77 ± 0.30	5.60 ± 0.56	1.16 ± 0.31	(0.08, 2.34)	0.0002
Diagnostic studies	6.89 ± 0.29	5.62 ± 0.08	1.26 ± 0.29	(0.68, 1.84)	<0.0001
Disposition	6.54 ± 0.31	5.52 ± 0.09	1.01 ± 0.31	(0.39, 1.63)	0.0015
Emergency stabilization	6.22 ± 0.30	5.51 ± 0.08	0.70 ± 0.30	(0.10, 1.30)	0.0212
History and physical	6.95 ± 0.33	5.72 ± 0.08	1.23 ± 0.33	(0.57, 1.89)	0.0003
Multi-tasking	6.80 ± 0.33	5.48 ± 0.08	1.31 ± 0.33	(0.65, 1.97)	0.0001
Team management	5.99 ± 0.31	5.47 ± 0.10	0.52 ± 0.30	(-0.08, 1.13)	0.0902

Table 2. Comparison of post graduate year 1 self-rating to attending rating.

Evaluation construct	Self	Rater(s)	Estimated difference ± standard error	95% CL	*p*
Communication	6.16 ± 0.60	4.14 ± 0.21	2.01 ± 0.60	(0.81, 3.21)	0.001
Diagnosis	5.17 ± 0.56	3.95 ± 0.16	1.21 ± 0.57	(0.09, 2.34)	0.0348
Diagnostic studies	6.56 ± 0.54	4.03 ± 0.17	2.52 ± 0.55	(1.45, 3.60)	<0.0001
Disposition	5.35 ± 0.58	3.92 ± 0.19	1.42 ± 0.58	(0.27, 2.58)	0.0152
Emergency stabilization	4.95 ± 0.56	3.69 ± 0.18	1.25 ± 0.56	(0.14, 2.36)	0.0265
History & physical	6.56 ± 0.61	4.12 ± 0.18	2.44 ± 0.62	(1.21, 3.66)	0.0001
Multi-tasking	5.17 ± 0.61	3.77 ± 0.17	1.39 ± 0.62	(0.17, 2.62)	0.0254
Team management	4.92 ± 0.57	4.07 ± 0.21	0.84 ± 0.57	(-0.27, 1.97)	0.1392

remains subject to these considerations. This suggests that educators must be cognizant of residents' self-assessments when formulating and delivering feedback. Our subgroup analysis included small sample sizes; more work with larger sample sizes is necessary to determine if program year does indeed have an effect on agreement between resident and attending assessment. Within this context, our data showed that differences between self-assessment and attending assessment may be affected by program year. Unlike the results for PGY 1 through 3, self and attending ratings for PGY 4 did not differ significantly on any sub-competency. These results suggest that in PGY 4, self and attending ratings

Table 3. Comparison of post graduate year 3 self-rating to attending rating.

Evaluation construct	Self	Rater(s)	Estimated difference ± standard error	95% CL	*p*
Communication	6.84 ± 0.55	5.79 ± 0.18	1.05 ± 0.55	(-0.03, 2.13)	0.0573
Diagnosis	7.50 ± 0.51	5.91 ± 0.14	1.59 ± 0.52	(0.56, 2.61)	0.0024
Diagnostic studies	7.17 ± 0.49	5.98 ± 0.15	1.19 ± 0.49	(0.21, 2.16)	0.0165
Disposition	6.51 ± 0.53	5.87 ± 0.16	0.63 ± 0.53	(-0.40, 1.68)	0.2291
Emergency stabilization	6.99 ± 0.51	5.89 ± 0.16	1.10 ± 0.51	(0.10, 2.11)	0.0311
History & physical	7.66 ± 0.55	6.08 ± 0.15	1.58 ± 0.56	(0.47, 2.68)	0.0053
Multi-tasking	7.49 ± 0.55	5.72 ± 0.15	1.76 ± 0.56	(0.65, 2.87)	0.0019
Team management	6.31 ± 0.52	5.68 ± 0.18	0.62 ± 0.51	(-0.39, 1.64)	0.2272

Table 4. Comparison of post graduate year 2 self-rating to attending rating.

Evaluation construct	Self	Rater(s)	Estimated difference ± standard error	95% CL	*P*
Communication	6.14 ± 0.67	5.07 ± 0.20	1.07 ± 0.67	(-0.24, 2.39)	0.1113
Diagnosis	6.70 ± 0.63	5.19 ± 0.15	1.50 ± 0.63	(0.25, 2.75)	0.0181
Diagnostic studies	6.43 ± 0.60	5.15 ± 0.16	1.28 ± 0.60	(0.09, 2.46)	0.0346
Disposition	6.91 ± 0.65	5.07 ± 0.18	1.83 ± 0.64	(0.56, 3.11)	0.0048
Emergency stabilization	5.87 ± 0.62	5.07 ± 0.17	0.79 ± 0.62	(-0.43, 2.02)	0.2043
History & physical	5.92 ± 0.68	5.20 ± 0.17	0.72 ± 0.68	(-0.62, 2.07)	0.2934
Multi-tasking	6.19 ± 0.68	5.10 ± 0.16	1.08 ± 0.68	(-0.26, 2.44)	0.1153
Team management	6.12 ± 0.64	5.03 ± 0.20	1.09 ± 0.63	(-0.15, 2.33)	0.0863

Table 5. Comparison of post graduate year 4 self-rating to attending rating.

Evaluation construct	Self	Rater(s)	Estimated difference ± standard error	95% CL	*P*
Communication	7.58 ± 0.78	7.15 ± 0.19	0.43 ± 0.77	(-1.09, 1.95)	0.58
Diagnosis	7.71 ± 0.72	7.35 ± 0.15	0.35 ± 0.73	(-1.08, 1.79)	0.626
Diagnostic studies	7.39 ± 0.69	7.34 ± 0.16	0.04 ± 0.69	(-1.31, 1.41)	0.9434
Disposition	7.38 ± 0.74	7.23 ± 0.18	0.15 ± 0.74	(-1.31, 1.62)	0.8399
Emergency stabilization	7.07 ± 0.72	7.40 ± 0.17	-0.33 ± 0.71	(-1.74, 1.08)	0.643
History & physical	7.67 ± 0.78	7.47 ± 0.16	0.19 ± 0.79	(-1.36, 1.75)	0.8054
Multi-tasking	8.33 ± 0.78	7.32 ± 0.16	1.01 ± 0.79	(-0.54, 2.57)	0.203
Team management	6.59 ± 0.73	7.06 ± 0.20	-0.46 ± 0.73	(-1.90, 0.97)	0.5271

converge and are quite similar. It is important to note that these *p* values were not adjusted for multiple comparisons and should therefore be interpreted as only part of further exploratory analyses.

Taking the results of our study into consideration, the finding that residents perceive themselves as more capable than they are rated by attendings would be relevant to discussions in Clinical Competency Committee (CCC) meetings. Residents' perception of their skills would be important in grading them on sub-competencies that deal with "practice-based learning and improvement." Although difficult to put into practice, perhaps resident self-evaluations should be included in their "residency portfolios" and compared to the CCC rating of that resident to ensure that as the resident moves through the program self-perception is not significantly different from that of his/her evaluators.

LIMITATIONS

The major limitation of this study is the small sample size of attendings and residents evaluated. Our self-assessment response rate was 70%. Self-assessment was not compulsory, as participation in research was voluntary per our HIC. It is not known if those who did not respond were different demographically or in PGY year, as the study personnel was blinded to the identities of the residents assessed. Another limitation is the varying levels of familiarity with the milestones among the residents and attendings surveyed. In addition, due to the nature of EM shift work, attendings have different frequencies of interactions with residents, which may introduce bias into their assessments. As the study was done

at an academic institution, some faculty members do have less clinical time in the department than others. Although faculty members were permitted to opt out of assessing a resident with whom they had limited experience, we acknowledge that the frequency and types of faculty-resident interactions assessed may vary widely. In addition, the residents studied were at different levels of their training; this may have influenced their self-ratings. For example, some of the residents surveyed were close to graduation, a circumstance that may inflate their self-assessments. Although our data suggest that PGY 4 residents' and attending evaluations may be similar, our interpretation is limited by the small number of representatives in each class; thus, more investigation is required to determine if there is a difference between classes in their ability to self-assess accurately. A larger sample of residents assessed may allow for more detailed sub-group analysis by PGY year. In addition, a larger sample size would also allow for more detailed analysis of high and low performers and their ability to self-assess, as had been demonstrated in the past. This study relies on the assumption that attending ratings are more accurate than resident self-rating, the validity of which may need further investigation.[4,7] Perhaps most importantly, milestones are a relatively new assessment tool with very few studies evaluating their validity.[1,3,9]

CONCLUSION

Residents over-estimated their abilities in each of eight sub-competencies assessed. This underscores the importance of feedback and assessment transparency. More attention needs to be paid to methods by which residency leadership can make residents' clinical ability self-perception more congruent with that of their teachers and evaluators.

Address for Correspondence: Katja Goldflam, MD, Yale University School of Medicine, 464 Congress Ave. Suite 260, New Haven, CT 06511. Email: katja.goldflam@yale.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Beeson MS, Carter WA, Christopher TA, et al. The development of the emergency medicine milestones. *Acad Emerg Med.* 2013;20(7):724-9.
2. Nasca TJ, Philibert I, Brigham T, et al. The next accreditation system rationale and benefits. *N Engl J Med.* 2012;366(11):1051–6.
3. Korte RC, Beeson MS, Russ CM, et al. Emergency Medicine Milestones Working Group, Reisdorff EJ. The emergency medicine milestones: a validation study. *Acad Emerg Med*. 2013;20(7):730–5.
4. Veloski J, Boex JR, Grasberger MJ, et al. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach*. 2006;28:117–28.
5. Sargeant J, Armson H, Chesluk B, et al. The processes and dimensions of informed self-assessment. *Acad Med*. 2010;85:1212–20.
6. Bounds R, Bush C, Aghera A, et al. MERC at CORD Feedback Study Group. Emergency medicine residents' self-assessments play a critical role when receiving feedback. *Acad Emerg Med*. 2013;20(10):1055-61.
7. Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence. *JAMA*. 2006;296:1094–102.
8. Eva KW, Armson H, Holmboe E, et al. Factors influencing the responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract*. 2012;17:15–26.
9. Peck TC, Dubosh N, Rosen C, et al. Practicing emergency physicians report performing well on most emergency medicine milestones. *J Emerg Med.* 2014;47(4):432-40.

Integration of a Blog into an Emergency Medicine Residency Curriculum

Jay Khadpe, MD
James Willis, MD
Mark A. Silverberg, MD
Andrew Grock, MD
Teresa Smith, MD

SUNY Downstate Medical Center, Department of Emergency Medicine, Brooklyn, New York

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted April 27, 2015; Revision received August 12, 2015; Accepted August 25, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.27199
[West J Emerg Med. 2015;16(6):936–937.]

BACKGROUND

Technologies and techniques for knowledge translation are rapidly evolving and there is a need for graduate medical education (GME) curricula to keep up with these advances to reach our learners in an effective manner. Technologies such as blogs, microblogs, wikis, podcasts, and vodcasts have the potential to expand upon the current didactic models by adding dimensions and engaging learners in modalities not previously available.[1]

Emergency medicine (EM) has been at the forefront in adopting social media in the pursuit of knowledge and collaboration.[2,3] In order to advise residencies on how to use these new technologies, the Council of EM Residency Directors (CORD) Social Media Task Force published its guidelines, best practices, and recommendations for integrating social media into EM residency programs laying the foundation for EM residency-based social media activities.[4] In March 2012, "The Original Kings of County" (TOKC) blog was launched in an effort to integrate social media into the SUNY Downstate/Kings County Hospital EM Residency Program (Figure). The blog represents an early adapter of these efforts and uniquely applies them at the GME level to improve learner engagement with the EM residency curriculum.

EDUCATIONAL OBJECTIVES

1. To increase resident engagement in their didactic curriculum through use of a residency blog.
2. To develop residents' skills with respect to education and scholarship through authorship for a residency blog.

CURRICULAR DESIGN

The TOKC blog was implemented to create an online hub for the integration of social media into the residency curriculum

Figure. The front page of The Original Kings of County blog.

at the SUNY Downstate / Kings County Hospital EM residency. Three overarching goals drive the content for the blog. The first is to post educational content that mirrors didactics already occurring within the program to reinforce the material and share it with residents who are unable to attend these activities. Examples include our Morning Reports that are typically brief case discussions with clinical pearls written by senior residents and a "Wednesday Wrap-up" that summarizes learning points and resources related to topics discussed during weekly conference didactics. By posting these resources on the blog, the impact of these didactics is magnified and enables all learners to benefit.

The second goal is to engage residents through their own authorship of "featured" blog posts. This activity often

coincides with the residents' development of an academic niche as they hone their skills as academic authors. We have used a mixed editorial process with respect to these posts decided between the author and blog editors. In some cases, a faculty editor aids with post development prior to publication; and in other cases, the resident will post directly to the blog and post-publication review will take place through the comment section. By allowing for both editorial formats, residents were more enthusiastic about authoring and contributing to the blog.

The third goal is intended to attract and encourage participation in the blog. Clinical cases are posted for residents to interpret and discuss. The resident who submits the most accurate and inclusive interpretation, as judged by the author and faculty editor, receives a small prize, such as a gift card. For example, residents rotating in the coronary care unit are asked to submit an electrocardiogram, which is then reviewed by a resident author and faculty editor prior to posting. These competitions are an attempt to embrace the competitive spirit among residents together with a monetary award to drive learners to the blog who might otherwise not participate; and once at the site, engage them with the other available content. Overall, this three-prong approach creates a comprehensive online didactic presence that embraces the use of social media to promote learning.

IMPACT/EFFECTIVENESS

Since launching, TOKC has generated over 600 posts by more than 20 resident and faculty authors as of April 2015. It currently receives more than 100 page views per day. This provides our program a platform to share their scholarship with a local, national, and international community. Alumni authors of TOKC have gone on to blog for internationally recognized academic EM blogs, such as *Academic Life in EM*, and to lecture at national conferences on how to use social media for medical education, demonstrating how participation in a residency-based blog can aid in the development of an academic niche. Additionally, TOKC was highlighted in the article, "Integration of Social Media in EM Residency Curriculum," by Scott et al. published in *Annals of EM* as a model for integrating a blog into an EM residency program.[5]

Address for Correspondence: Jay Khadpe, MD, SUNY Downstate Medical Center, 451 Clarkson Avenue, Box 1228, Brooklyn, NY, 11203. Email: jay.khadpe@downstate.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Boulos MN, Maramba I, Wheeler S. Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Med Educ.* 2006;6:41.
2. Cadogan M, Thoma B, Chan TM, et al. Free Open Access Meducation (FOAM): the rise of emergency medicine and critical care blogs and podcasts (2002-2013). *Emerg Med J.* 2014;31(e1):e76-77.
3. Genes N. "Getting Social." *Emergency physicians monthly*. June 2011.
4. Pillow MT, Hopson L, Bond M, et al. Social media guidelines and best practices: recommendations from the council of residency directors social media task force. *West J Emerg Med.* 2014;15(1):26-30.
5. Scott KR, Hsu CH, Johnson NJ, et al. Integration of social media in emergency medicine residency curriculum. *Ann Emerg Med.* 2014;64(4):396-404.

Ultrasound Training in the Emergency Medicine Clerkship

Mark Favot, MD*
Cheryl Courage, MA*
Jacob Mantouffel, MD†
David Amponsah, MD†

*Wayne State University School of Medicine, Department of Emergency Medicine, Detroit, Michigan
†Henry Ford Health System, Department of Emergency Medicine, Detroit, Michigan

Section Editor: David Wald, MD
Submission history: Submitted April 30, 2015; Revision received September 7, 2015; Accepted September 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27290

Introduction: The curriculum in most emergency medicine (EM) clerkships includes very little formalized training in point-of-care ultrasound. Medical schools have begun to implement ultrasound training in the pre-clinical curriculum, and the EM clerkship is an appropriate place to build upon this training. The objectives are (1) to evaluate the effectiveness of implementing a focused ultrasound curriculum within an established EM clerkship and (2) to obtain feedback from medical students regarding the program.

Methods: We conducted a prospective cohort study of medical students during an EM clerkship year from July 1, 2011, to June 30, 2012. Participants included fourth-year medical students (n=45) enrolled in the EM clerkship at our institution. The students underwent a structured program focused on the focused assessment with sonography for trauma exam and ultrasound-guided vascular access. At the conclusion of the rotation, they took a 10-item multiple choice test assessing knowledge and image interpretation skills. A cohort of EM residents (n=20) also took the multiple choice test but did not participate in the training with the students. We used an independent samples *t*-test to examine differences in test scores between the groups.

Results: The medical students in the ultrasound training program scored significantly higher on the multiple-choice test than the EM residents, $t(63)=2.3$, $p<0.05$. The feedback from the students indicated that 82.8% were using ultrasound on their current rotations and the majority (55.2%) felt that the one-on-one scanning shift was the most valuable aspect of the curriculum.

Discussion: Our study demonstrates support for an ultrasound training program for medical students in the EM clerkship. After completing the training, students were able to perform similarly to EM residents on a knowledge-based exam. [West J Emerg Med. 2015;16(6):938–942.]

INTRODUCTION

Often, the only opportunity medical students have to spend a significant amount of time in the emergency department (ED), caring for acutely ill, undifferentiated patients is during the emergency medicine (EM) clerkship,[1] which typically take place exclusively during the fourth-year of medical school. In addition to EM, medical students in EM clerkships will do their residency training in a variety of specialities such as surgery, internal medicine, and obstetrics/gynecology, each of which uses point-of-care ultrasound (POCUS) in their practice.[2-5] However, given the current structure of the clinical clerkship curriculum at the majority of medical schools, students are not routinely exposed to POCUS.[6] In fact, the majority of medical students receive no formal education in ultrasound during medical school,[7,8] and in the instances when they do receive training, it is unclear if they develop competency.[9]

The Accreditation Council for Graduate Medical Education (ACGME) guidelines for specialty training list POCUS applications as a requirement of nearly every

residency program.[2-5,10] Since many residents struggle to gain competence in POCUS, we believe that the earlier and more often training is implemented into the undergraduate medical education curricula, the more likely these students will be proficient upon completion of their residency training. A study of first-year medical students demonstrated that after training they could assess the abdominal aorta of healthy volunteers to the same standard achieved by professional sonographers.[11] These students received just four hours of formal instruction from a single physician in a small group setting, giving encouragement to educators that effective ultrasound education to students does not require a burdensome commitment of resources.

There are many factors outside of the students' control that interfere with the opportunity to gain experience with POCUS in the ED, including the fast pace of care, high volume and higher acuity of patients. Given these barriers, it is common for students to complete an EM rotation with limited experience in this rapidly growing field.[12,13] It is therefore imperative that medical students gain experience with this modality during formalized periods of instruction, in addition to exposure during their ED shifts. The EM clerkship provides an excellent opportunity to fill this gap in the undergraduate medical curriculum. The purposes of this study were (1) to evaluate the effectiveness of implementing a focused ultrasound curriculum within an established EM clerkship and (2) to obtain feedback from medical students regarding the program.

METHODS

We conducted a prospective static-group comparison study of medical students during an EM clerkship conducted from July 1, 2011 to June 30, 2012.

Setting

The setting was a large, urban, academic medical center, a Level 1 trauma center with an annual ED census of approximately 96,000 patients. The hospital has a large EM residency program (n=42 categorical residents), as well as combined residency programs in EM/Internal Medicine and EM/Internal Medicine/Critical Care Medicine. The department also has a fellowship in emergency ultrasound. We obtained institutional review board approval, and a waiver of informed consent was granted as the project was a part of an educational curriculum and participation was voluntary.

Participants

Participants included fourth-year medical students (n=45) in the EM clerkship during the 2011-2012 academic year. Participation was optional, and lack of participation had no impact on clerkship grade. Students were excluded from participating if they were unable to attend the introductory training session. Figure 1 contains information regarding the chosen specialties of the students. Approximately half of the students identified themselves as pursuing a career in EM,

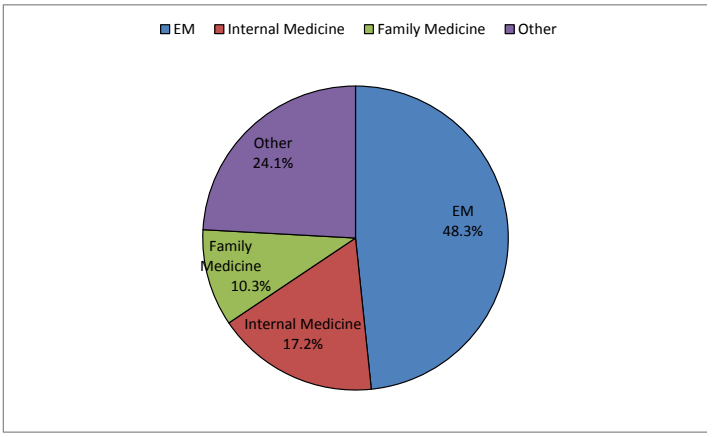


Figure 1. Medical students' identified specialty of interest.
EM, emergency medicine

while internal medicine and family medicine were the next most popular choices.

Due primarily to time constraints (one year for data collection) and the small number of rotating medical students, a control group consisting of medical students was not feasible; therefore, we used a non-randomized sample of EM residents (n=20) to serve as a comparison group. Members of the comparison group did not partake in any of the curricular components of the program, but each of them had received formal training in ultrasound during their residency, including an introductory course in emergency ultrasound during intern year. The concepts taught to the students during the didactic and hands-on components of the curriculum were also taught to the residents during their intern year as part of the residency curriculum.

Methods of Instruction

A POCUS curriculum for EM clerkship students was developed by the faculty of the Emergency ultrasound fellowship program. The curriculum centered on the focused assessment with sonography for trauma (FAST) exam and ultrasound-guided vascular access. We felt that these applications were most relevant to the student due to their frequent use in clinical practice. Methods of instruction included didactic presentations, hands-on ultrasound scanning of live human models and tissue phantoms, and a dedicated ultrasound scanning shift working one-on-one with the ultrasound fellow in the ED.

After orientation on the first day of the clerkship, participants attended a two-hour instructional session on the FAST exam and ultrasound-guided vascular access techniques. This introductory session was led by the fellow and incorporated didactic lecture and hands-on practice in our institution's simulation center. The didactic session lasted approximately 75 minutes, with 50 minutes devoted to the FAST exam and 25 minutes to vascular access. The ultrasound machines used in the simulation center were a GE

Logiq E and a GE Venue 40 (Milwaukee, WI). Each student was required to perform a minimum of two ultrasound-guided vascular access procedures, one on a peripheral vessel phantom, and one on a central venous access mannequin, as well as two FAST exams.

Following this session, students were assigned a dedicated ultrasound shift in the ED to reinforce the techniques learned in the simulation center. During these shifts all students were allowed the opportunity to perform multiple FAST examinations and attempt ultrasound-guided, peripheral venous access procedures on ED patients. Throughout the clerkship, the students were able to perform FAST exams on all trauma activations. Additionally, students had the opportunity to place both peripheral and central lines in the course of their clerkship, many of which were done with ultrasound guidance, thus allowing further opportunity for skill refinement.

Impact/Effectiveness

Assessment

A 10-item multiple choice test was developed to test students' knowledge on the FAST exam and ultrasound-guided vascular access. Items and content on the test had been used previously during the pre-clinical POCUS training that the emergency ultrasound fellowship faculty has led for the past six years.[14,15] The test was administered to the students individually during the last week of the clerkship under the supervision of the principle investigator. The EM residents were given the same test during the mid-point of the study period. Residents were approached by the investigators during a regularly scheduled conference and asked to participate. Participation by residents was voluntary and did not impact status within the program. The residents completed the test simultaneously under the supervision of the principle investigator. To ensure responses were anonymous, we collected only residents' level of training and test responses.

One month after the conclusion of the clerkship, each student was contacted by the primary investigator by email and asked to respond to a brief, seven-item survey regarding their ultrasound training experience. Survey questions are presented in Table 1.

Statistical Analysis

We used an independent samples t-test to examine differences in test scores between students and residents. We performed all analyses with IBM SPSS 20.

RESULTS

The mean test score for the medical student participants in the ultrasound curriculum was 8.3 (SD 1.2) and the mean test score for the residents in the comparison group was 7.6 (SD 1.1). This difference in score between the two groups was statistically significant (p=0.026). Table 2 and Figure 2 show additional details of the test scores for the medical students and the residents.

Table 1. Student survey responses regarding the point-of-care ultrasound training program.

Question	Response
What specialty are you applying to?	EM (n=14), Internal Medicine (n=5), Family Medicine (n=3), Other (n=7): Surgery, Radiology, Neurology, Anesthesiology and Ophthalmology. See Figure 1.
Are you utilizing US on current rotations?	Yes (82.8%)
Are you utilizing US on Non-EM rotations?	Yes (58.6%)
Do you see yourself utilizing US in residency?	Yes (93.1%)
What did you find most useful about curriculum?	One-on-one instruction (55.2%)
What other applications would you have liked to learn about?	Echocardiography (24.1%); MSK (10.3%); Abscess (6.9%); DVT (6.9%)
How can we improve the curriculum?	Additional scanning shifts (51.7%); More evaluation of patients with pathology (24.1%)

EM, emergency medicine; *US,* ultrasound; *MSK,* musculoskeletal; *DVT,* deep vein thrombosis

The results from the feedback survey administered to the students are presented in Table 1. The survey response rate was 64% (n=29). Most of the students stated that they were using ultrasound on current or subsequent rotations (82.8%) and planned to use it during their residency (93.1%). A majority of the students replied that they valued the one-on-one instruction (55.2%) and would choose to add more than one scanning shift (51.7%) when asked how the curriculum could be improved. Of different ultrasound applications, students most wanted additional experience with bedside echocardiography (24.1%).

DISCUSSION

The current study demonstrates that medical students are able to perform similarly to EM residents on a multiple-choice examination following the completion of an EM clerkship with a focused POC ultrasound curriculum.

Few studies of medical students performing ultrasound have demonstrated competence either in terms of knowledge or in skills such as image acquisition or interpretation. In a study with first-year medical students incorporating an objective, knowledge-based ultrasound test and a practical hands-on examination, investigators compared students' performance pre- and post-educational intervention, randomizing students into two groups, "early" and "late" intervention.[10] Both groups demonstrated improved

Table 2. Multiple-choice ultrasound knowledge test scores.

Group	Mean* (SD)	Median	Minimum	Maximum
Medical students (n=45)	8.3 (1.2)	8.0	6.0	10.0
EM residents (n=20)	7.6 (1.1)	8.0	5.0	9.0

SD, standard deviation; *EM,* emergency medicine
*Mean scores were significantly different between students and residents t(63)=2.3, p=0.026.

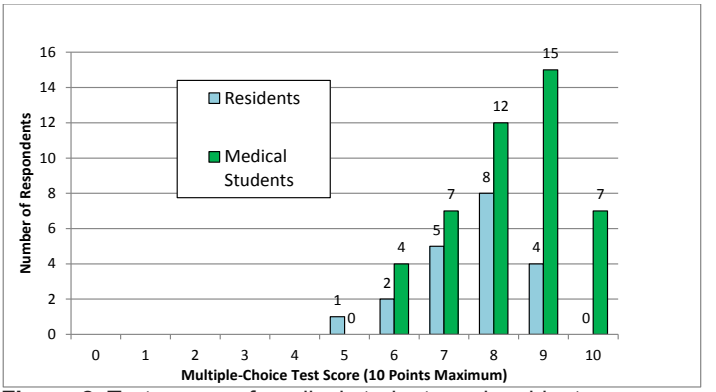


Figure 2. Test scores of medical students and residents.

performance after the intervention, yet they did not compare performance with a group of subjects who had already achieved some level of competence in POCUS. The pre- and post-test method of determining whether medical students gained knowledge is used frequently in the literature in training undergraduate medical students in POCUS.[16,17]

Another education research method frequently encountered in the literature is the assessment of the confidence and satisfaction of the learners after they have completed the curriculum.[16,18,19] We feel that the follow-up survey responses we obtained offer particularly high-yield insights to educators looking to introduce POCUS curricula into existing EM clerkships. The feedback not only shows high levels of satisfaction but also offers potential course improvements, such as including more than one scanning shift and adding other applications such as POC echocardiography to the program. Thus, clerkship directors at programs with core ultrasound faculty could replicate our curriculum and enhance it by following some of the recommendations made by the students.

LIMITATIONS

Our study did suffer from several limitations. The sample sizes were relatively small, and we were unable to include all of the students in the clerkship into the study because of logistical issues related to differing rotation start dates for students from various medical schools. Additionally, approximately two-thirds of the students had received previous ultrasound training.[15] Thus, it is possible that our cohort of students scored better on the test than residents as a result of their previous ultrasound training, not because of the effectiveness of the ultrasound curriculum. Also, the qualitative data we received from the feedback survey had a moderate response rate. The potential exclusion of students who did not have a favorable opinion of the program could have skewed our results. Lastly, the residents who took the test to serve as the comparison group also suffered from a low response rate, as less than 50% of the residents submitted a completed version of the test to the investigators.

It is also worth noting that because of the limitations of designing research within the constraints of an established clerkship, we did not have a true control group of medical students who were tested but not exposed to the educational intervention. Lastly, while our data demonstrate that medical students were able to perform at a high level on a knowledge-based examination, the higher scores, while statistically significant, amounts to on average just one question on a 10-question exam. A clinical significance is likely negligible.

CONCLUSION

Our study demonstrates that with a dedicated, integrated curriculum in POC ultrasound fourth-year medical students on an EM clerkship are able to perform similarly to EM residents on a knowledge-based exam. Further, our research shows that medical students find POCUS training to be a useful adjunct to the EM clerkship curriculum and that they feel the skills they acquire during this time will serve to benefit them on future clerkships and during residency training. Educators who are attempting to develop POCUS curricula for their existing EM clerkships would be wise to heed the advice of the students and include more than one scanning shift and also add POC echocardiography to the curriculum.

Address for Correspondence: Mark Favot, MD, Wayne State University, 6071 W. Outer Dr., Lourdes 447, Detroit, MI 48235. Email: mfavot@med.wayne.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Manthey DE, Ander DS, Gordon DC, et al. Emergency medicine

clerkship curriculum: an update and revision. *Acad Emerg Med*. 2010;17(6):638-43.

2. Emergency Medicine Guidelines. *Accreditation Council for Graduate Medical Education (ACGME)* Available at: http://www.acgme.org/acgmeweb/tabid/292/ProgramandInstitutionalGuidelines/HospitalBasedAccreditation/EmergencyMedicine/.EmergencyMedicineGuidelines.aspx. Accessed Aug 13, 2013.
3. ACGME Program Requirements for Graduate Medical Education in Obstetrics and Gynecology. *Accreditation Council for Graduate Medical Education (ACGME)*. Available at: http://www.acgme.org/acWebsite/downloads/RRC_progReq/220obstetricsandgynecology01012008.pdf. Accessed Aug 15, 2013.
4. Program Requirements for Graduate Medical Education in Diagnostic Radiology. *Accreditation Council for Graduate Medical Education (ACGME) website*. Available at: https://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/420_diagnostic_radiology_07012008.pdf. Accessed Aug 15, 2013.
5. Akhtar S, Theodoro D, Gaspari R, et al. Resident training in emergency ultrasound: Consensus recommendations from the 2008 Council of Emergency Medicine Residency Directors Conference. *Acad Emerg Med.* 2009;16.s2:S32-S6.
6. Kman NE, Bernard AW, Martin DR, et al. Advanced topics in emergency medicine: Curriculum development and initial evaluation. *WestJEM.* 2011;12(4):543-50.
7. Lyss-Lerman P, Teherani A, Aagaard E, et al. What training is needed in the fourth year of medical school? Views of residency program directors. *Acad Med*. 2009;84(7):823-9.
8. Hoppmann RA, Rao VV, Poston MB, et al. An integrated ultrasound curriculum (iUSC) for medical students: 4-year experience. *Crit Ultrasound J.* 2011;3(1):1-12.
9. Bahner DP, Adkins EJ, Nagel R, et al. Brightness mode quality ultrasound imaging examination technique (B-QUIET): Quantifying quality in ultrasound imaging. *J Ultrasound Med.* 2011;30(12):1649-55.
10. Jeppesen KM and Bahner DP. Teaching bedside sonography using peer mentoring. *J Ultrasound Med.* 2012;31:455-9.
11. Wong I, Jayatilleke T, Kendall R, et al. Feasibility of a Focused Ultrasound Training Programme for Medical Undergraduate Students. *Clin Teach.* 2011;8(1):3-7.
12. Coates WC. An educator's guide to teaching emergency medicine to medical students. *Acad Emerg Med.* 2004;11(3):300-6.
13. Rao S, Holsbeeck LV, Musial JL, et al. A pilot study of comprehensive ultrasound education at the Wayne State University School of Medicine. *J Ultrasound Med.* 2008;27(5):745-9.
14. Fox JC, Anderson CL, Ahmed SS, et al. Effect of a medical student emergency ultrasound clerkship on number of emergency department ultrasounds. *WestJEM.* 2010;11(1):31-4.
15. Rao S, Holsbeeck LV, Musial JL, et al. A pilot study of comprehensive ultrasound education at the Wayne State University School of Medicine. *J Ultrasound Med.* 2008;27(5):745-9.
16. Afonso N, Amponsah D, Yang J, et al. Adding new tools to the black bag—Introduction of ultrasound into the Physical Diagnosis course. *J Gen Intern Med.* 2010;25(11):1248-52.
17. Butter J, Grant TH, Egan M. Does ultrasound training boost Year 1 medical student competence and confidence when learning abdominal examination? *Med Educ.* 2007;41:843-8.
18. Fox JC, Cusick S, Scruggs W, et al. Educational Assessment of Medical Student Rotation in Emergency Ultrasound. *WestJEM.* 2007;8(3):84-7.
19. Keddis MT, Cullen MW, Reed DA, et al. . Effectiveness of an Ultrasound Training Module for Internal Medicine Residents. *BMC Med Educ.* 2011;11:75.
20. Arger PH, Schultz SM, Sehgal CM. Teaching medical students diagnostic sonography. *J Ultrasound Med.* 2005;24(10):1365-9.

Assessing EM Patient Safety and Quality Improvement Milestones Using a Novel Debate Format

Mira Mamtani, MD
Kevin R. Scott, MD
Francis J. DeRoos, MD
Lauren W. Conlon, MD

University of Pennsylvania Perelman School of Medicine, Department of Emergency Medicine, Philadelphia, Pennsylvania

Section Editor: Sally Santen, MD, PhD
Submission history: Submitted April 30, 2015; Revision received September 21, 2015; Accepted September 21, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27269

Graduate medical education is increasingly focused on patient safety and quality improvement; training programs must adapt their curriculum to address these changes. We propose a novel curriculum for emergency medicine (EM) residency training programs specifically addressing patient safety, systems-based management, and practice-based performance improvement, called "EM Debates." Following implementation of this educational curriculum, we performed a cross-sectional study to evaluate the curriculum through resident self-assessment. Additionally, a cross-sectional study to determine the ED clinical competency committee's (CCC) ability to assess residents on specific competencies was performed. Residents were overall very positive towards the implementation of the debates. Of those participating in a debate, 71% felt that it improved their individual performance within a specific topic, and 100% of those that led a debate felt that they could propose an evidence-based approach to a specific topic. The CCC found that it was easier to assess milestones in patient safety, systems-based management, and practice-based performance improvement (sub-competencies 16, 17, and 19) compared to prior to the implementation of the debates. The debates have been a helpful venue to teach EM residents about patient safety concepts, identifying medical errors, and process improvement. [West J Emerg Med. 2015;16(6):943–946.]

BACKGROUND

Educating the next generation of resident physicians includes not only specialty-specific content but also patient-centered training focusing on value-based, high-quality care.[1] In early 2000, the Institute of Medicine released their report, "To Err is Human," highlighting the prevalence of medical error and suboptimal care.[2] Since then, patient safety and quality improvement has become a more visible topic to hospital management, the public, and medical educators. In 2012 the Emergency Medicine (EM) Milestone Project was created by the American Board of EM (ABEM) and the Accreditation Council for Graduate Medical Education (ACGME). The EM Milestone Project is a framework of assessing competencies within several domains of EM skill sets, including sub-competencies in patient safety, systems-based management, and practice-based performance improvement (sub-competencies 16, 17, and 19).[3] The ACGME further emphasized a learning environment focusing on patient safety and quality improvement by implementing the Next Accreditation System (NAS) in 2013.[1]

Educating resident physicians on quality improvement (QI) and patient safety has the potential to reduce errors and improve patient outcomes. In 2007, a systematic review concluded that teaching quality improvement to clinicians improved their knowledge and confidence.[4] Several studies have shown that participation in a QI curriculum resulted in significant improvement in processes of care.[5-9] A separate study published in 2010 reviewed barriers of effectively teaching quality improvement and patient safety to learners, which emphasized achieving the appropriate balance of didactic and experiential learning, as well as scheduling the curriculum amidst preexisting rotations.[10]

The authors describe a curriculum for EM residents, "EM Debates," which combine didactic and experiential learning during resident conference as a method to teach

and assess specific sub-competencies within the EM milestones. We sought to evaluate the effectiveness of this curriculum by surveying the residents' perception of the "EM Debates," and to determine whether the CCC could more easily assess specific patient safety and quality improvement milestones.

Objectives

The proposed curriculum, "EM Debates," involves a mock clinical case that is debated by two teams: a senior-level resident and attending team debating against another senior-level resident and attending team. A moderator introduces the topic and poses questions to the debaters as well as the audience participants. The teams have 15 minutes each to present their opposing views for the diagnosis, treatment, and/or disposition of commonly encountered emergency department (ED) presentations using the best available evidence. Following this, the audience participants have 15 minutes to discuss the opposing viewpoints and to determine if a consensus can be reached.

For example, a recent "EM Debate" involved the treatment of patients with atrial fibrillation. A case of a patient with acute onset atrial fibrillation who had no comorbidities and no symptoms was presented. Articles were reviewed arguing that patients should be cardioverted in the ED and discharged home. The opposing team reviewed articles arguing that patients should be anticoagulated and placed in the hospital for delayed cardioversion. Following the discussion, the conference attendees (residents, faculty, and nurses) debated the relative merits of each side and eventually came to a consensus. The residents who presented the debate then worked with key QI faculty members within our department and the department of cardiology to create a pathway for patients with atrial fibrillation.

The objectives of this curriculum are separated for those who participate in an "EM Debate," lead a debate, and develop a protocol following the completion of the debate. The goals for the debate leader include developing an in-depth understanding of a controversial topic, critically analyzing current literature with a faculty mentor, and creating a persuasive argument to teach the participants about managing a specific disease process. The debate leaders could be assessed on leading team reflections to improve ED performance and demonstrating evidence-based information retrieval mastery. These assessments would fall within patient safety sub-competency 16 and practice-based performance improvement sub-competency 19.

The goals for the debate participants, those residents in the audience, include describing the best available evidence or controversies surrounding a specific topic. Additionally, participants will appraise the value of the leader's presentations and choose a management strategy they will adopt for a specific clinical question. The debate participants could be assessed on their ability to describe patient safety concepts (sub-competency 16, level 3) and identifying situations when breakdown in teamwork or communication may contribute to medical error (sub-competency 16, level 4). Additionally, the participants could be assessed on the ability to call effectively on other resources in the system (sub-competency 17, level 3), and the ability to critically appraise literature and apply evidence-based medicine (sub-competency 19, level 3).

The goals for the debate leaders who participate in protocol development following a debate include synthesizing the current literature and the feedback from the participants during the debate to propose a clinical pathway for departmental practice. By completing the pathway, the senior level resident will have addressed sub-competency 19 at level 4; specifically, they would have participated in a process improvement plan to optimize ED performance and applied performance improvement methodologies. Furthermore, they will have addressed sub-competency 17 at level 4; specifically, they would have participated in processes and logistics to improve patient flow and decrease turnaround time.

Curricular Design

This curriculum was guided by the six-step model for medical curriculum development. The six-step approach highlights a process of general needs assessment, targeted needs assessment, goals and objectives, educational strategies, implementation, and evaluation and feedback.[11]

Implementation of the EM milestones highlighted the general needs assessment within graduate medical education on teaching patient safety and QI. Within our program we targeted controversial topics that were seen as patient safety issues. We felt our learners would benefit from active participation in evidence-based discussions regarding clinical management strategies and for our senior residents to develop protocols with department leadership. The goals and objectives outlined above directly address sub-competencies within the EM milestones. Using a debate format we were able to engage learners at various levels and promote discussion. Implementation during regular EM conference time maximized resident and faculty participation. Additionally, securing continuing medical education credit under the patient safety designation for the "EM debates" offered added incentive for faculty attendance. Evaluation and feedback of our curriculum are ongoing and will be discussed later in this manuscript.

IMPACT/EFFECTIVENESS

The "EM Debates" were implemented two years ago at our institution. We conducted a cross-sectional study to evaluate the effectiveness of this new curriculum. The survey was distributed to the EM residents at a single site to determine their perceptions of the debates using a Likert

scale. The responses were anonymous and participation was voluntary. Questions were divided into sections based on whether the respondent participated in a debate, led a debate, or were working on developing a pathway for the department after completion of a debate. The questions are linked to specific milestones. For example, one question: "After participating in a debate, I feel I have improved my individual performance on a specific topic by critically appraising scientific literature and applying evidence-based medicine," is linked to sub-competency 16, level 4 of the EM milestones.

We also conducted a cross-sectional study to determine the clinical competency committee's (CCC's) ability to assess residents on specific competencies. We sent a separate survey to CCC members to assess whether it was easier to assess these competencies after the implementation of the "EM Debates." Again, the responses were anonymous and participation was voluntary.

The institutional review board at our site granted exempted approval to this study. We created and distributed the survey using the online survey tool SurveyMonkey.© Survey responses were collected and compiled. We used descriptive quantitative and qualitative statistics to assess survey responses.

RESULTS

Quantitative

Residents

The survey was sent to 42 residents; of those, 30 residents responded (71% response rate). Seventy-one percent (71%) of these residents agreed or strongly agreed that they have improved their individual performance on a specific topic by critically appraising scientific literature and applying evidence-based medicine, which linked to sub-competency 19, level 3. Sixty-eight percent (68%) agreed or strongly agreed they could describe patient safety concepts, like the "Swiss cheese" or "near miss" model, which linked to sub-competency 16, level 3.

Approximately a third of the residents have reported that they have led a debate. Of those, 100% felt that they agreed or strongly agreed that they can propose an evidence-based approach to a specific topic, and 91% agreed or strongly agreed that they have analyzed or worked on improving ED performance, correlating to sub-competency 16 (level 4) and 19 (level 4), respectively.

Of those that led a debate, approximately half responded that they have worked on a protocol for the department. As only our third-year residents have led a debate, and we have completed half of them this year, this is an accurate representation. Of these residents, 100% agree or strongly agree that they have applied performance improvement methodologies and have analyzed processes and logistics to improve patient flow and turnaround time, corresponding to sub-competency 19 (level 4) and 17 (level 4), respectively.

CCC

The CCC felt that it was easier to assess sub-competency 16 levels 3 and 4, sub-competency 17 levels 3 and 4, and sub-competency 19 levels 3 and 4 as compared to prior to the implementation of the debates.

Qualitative

Respondents were overall very positive towards the implementation of the debates. A majority of the residents commented on the engaging aspect of discussing opposing management strategies of controversial topics, while others valued the detailed literature review. Some respondents, particularly junior residents, would appreciate more background information at the beginning of the debate to understand why the clinical case can have various management options. Additionally, multiple respondents felt that additional time devoted to the "EM Debates" during conference would be helpful.

LIMITATIONS

While the evaluation of the curriculum "EM Debates" indicated that the residents enjoyed it and, in certain instances, the pathway development changed clinical practice, it does not reveal whether it changed the residents' behavior. Future studies could be performed using direct observation during clinical practice to determine whether residents' behavior has changed as a result of this curriculum.

Furthermore, assessment of the debate leaders and the audience participants relies on subjective data. The assessment of the residents could be more robust if faculty with expertise in the area reviewed videotapes of the "EM Debates" to characterize the strength and validity of the debate leaders' argument and the audience participants' involvement in the discussion. A potential method to assess participants could include a post-debate examination to determine their level of understanding and ability to critically appraise the discussed literature. This would provide the CCC with more objective data of the residents' abilities.

This study was performed at a single site; further studies would need to be performed at additional sites to evaluate whether hospitals with different resources and cultures would find this as helpful.

DISCUSSION

Implementation of this curriculum is directly applicable and feasible among other EM programs. While faculty involvement could be a barrier, core faculty can be used as faculty mentors for the residents.

One challenge we noticed was the process of creating a protocol for the department is predicated on agreement when debating specific topics. If agreement is not reached, then the residents will delay creating a protocol until consensus within the department can be reached. Thus, specific milestones will not be assessed for those who have led a debate but

were unable to develop a protocol. For those residents, we found that assigning them to help with the creation of other protocols in which consensus was reached was a reasonable alternative. Furthermore, the success of pathway development is inherently dependent upon collegial interdepartmental relationships; inclusion of outside departments during the debates can help facilitate this.

While this is not a comprehensive quality and safety curriculum, it does include five of the eight Institute for Healthcare Improvement domains for health professional students: healthcare as a process, variation and measurement, collaboration, leading, following, and making changes, and developing new locally useful knowledge.[12]

The residents are generally very positive towards the "EM Debates," and it has been a helpful venue in teaching the residents about patient safety concepts, identifying medical errors, and process improvement. In addition, these debates have made it easier for the CCC to assess the sub-competencies practice-based performance improvement, patient safety, and systems-based management within the EM Milestone Project.

Address for Correspondence: Mira Mamtani, MD, Hospital of the University of Pennsylvania, 3400 Spruce St., Philadelphia, PA 19104. Email: mira.mamtani@uphs.upenn.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system rationale and benefits. *N Engl J Med*. 2012;366(11):1051-6.
2. Kohn LT, Corrigan J, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: *National Academies Press.* 2000.
3. Accreditation Council for Graduate Medical Education (ACGME). The Emergency Medicine Milestone Project. Available at: http://acgme.org/acgmeweb/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf. Accessed Apr 16, 2015.
4. Boonyasai RT, Windish DM, Chakraborti C, et al. Effectiveness of teaching quality improvement to clinicians: a systematic review. *JAMA*. 2007;298:1023-37.
5. Coleman MT, Nasraty S, Ostapchuk M, et al. Introducing practice-based learning and improvement ACGME core competencies into a family medicine residency curriculum. *Jt Comm J Qual Saf*. 2003;29:238-47.
6. Holmboe ES, Prince L, Green M. Teaching and improving quality of care in a primary care internal medicine residency clinic. *Acad Med*. 2005;80:571-7.
7. Mohr JJ, Randolph GD, Laughon MM, et al. Integrating improvement competencies into residency education: A pilot project from a pediatric continuity clinic. *Ambul Pediatr*. 2003;3:131-6
8. Oyler J, Vinci L, Arora V, et al. Teaching internal medicine residents quality improvement techniques using the ABIMs practice improvement modules. *J Gen Intern Med*. 2008;23:927-30.
9. Weingart SN,Tess A, Driver J, et al. Creating a quality improvement elective for medical house officers. *J Gen Intern Med*. 2004;19:861-7.
10. Wong BM, Etchells EE, Kuper A, et al. Teaching quality improvement and patient safety to trainees: a systematic review. *Acad Med*. 2010;85:1425-39.
11. Kern DE. Curriculum development for medical education: a six step approach. Baltimore, MD: Johns Hopkins University Press; 1998.
12. Ogrinc G, Headrick LA, Mutha S, et al. A framework for teaching medical students and residents about practice based learning and improvement, synthesized from a literature review. *Academic Medicine*. 2003;78(7):748-56.

Model for Developing Educational Research Productivity: The Medical Education Research Group

Marcia Perry, MD*
Laura Hopson, MD*
Joseph B. House, MD*
Jonathan P. Fischer, MPH‡
Suzanne Dooley-Hash, MD*
Samantha Hauff, MD*
Margaret S. Wolff, MD†
Cemal Sozener, MD*
Michele Nypaver, MD†
Joel Moll, MD*
Eve D. Losman, MD*
Michele Carney, MD†
Sally A. Santen, MD, PhD*

*University of Michigan, Department of Emergency Medicine, Ann Arbor, Michigan
†University of Michigan, Department of Emergency Medicine, Children's Emergency Services, Ann Arbor, Michigan
‡University of Michigan, Department of Public Health, Ann Arbor, Michigan

Section Editor: Jonathan Fisher, MD, MPH
Submission history: Submitted May 1, 2015; Revision received September 9, 2015; Accepted September 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27306

Introduction: Education research and scholarship are essential for promotion of faculty as well as dissemination of new educational practices. Educational faculty frequently spend the majority of their time on administrative and educational commitments and as a result educators often fall behind on scholarship and research. The objective of this educational advance is to promote scholarly productivity as a template for others to follow.

Methods: We formed the Medical Education Research Group (MERG) of education leaders from our emergency medicine residency, fellowship, and clerkship programs, as well as residents with a focus on education. First, we incorporated scholarship into the required activities of our education missions by evaluating the impact of programmatic changes and then submitting the curricula or process as peer-reviewed work. Second, we worked as a team, sharing projects that led to improved motivation, accountability, and work completion. Third, our monthly meetings served as brainstorming sessions for new projects, research skill building, and tracking work completion. Lastly, we incorporated a work-study graduate student to assist with basic but time-consuming tasks of completing manuscripts.

Results: The MERG group has been highly productive, achieving the following scholarship over a three-year period: 102 abstract presentations, 46 journal article publications, 13 MedEd Portal publications, 35 national didactic presentations and five faculty promotions to the next academic level.

Conclusion: An intentional focus on scholarship has led to a collaborative group of educators successfully improving their scholarship through team productivity, which ultimately leads to faculty promotions and dissemination of innovations in education. [West J Emerg Med. 2015;16(6):947–951.]

BACKGROUND

Education research and scholarship are essential for promotion of faculty as well as dissemination of new educational practices; however, it presents many

challenges. While there are a number of resources for faculty development,[1] for many faculty scholarship remains a daunting prospect.[2,3] Educational faculty spend the majority of their time on their education mission, leaving little time for scholarly pursuits. As a result, educators may lag on the scholarship essential for academic promotion. To combat this, creative tools are needed to promote scholarly activity for educational faculty.

Research literature on academic productivity indicates that institutional factors may play a larger role in determining research output than individual factors.[4] This insight has led to educational innovations about how to foster research environments that better promote productivity. Bland et al. found that factors promoting scholarship include (1) clear goals that serve as a coordinating function, (2) a distinctive culture of research emphasis with assertive participation, (3) frequent communication, (4) accessible resources, and (5) leadership with expertise and skill.[5] This work has served as a guide for research groups since its publication and has contributed to a growing body of education research literature. Questions remain regarding how to effectively implement the principles of effective research environments using these research group guidelines.

OBJECTIVES

As educational faculty face increasing demands on their time, it is imperative that new and creative models are developed to foster more productive research environments. The objective of this educational advance is to promote faculty scholarly productivity. We describe how the incorporation of research group guidelines to promote successful research through a multi-pronged approach led to scholarly productivity.[5] This method can provide a valuable template for other departments to follow.

EDUCATIONAL ADVANCE

The Medical Education Research Group (MERG) consists of faculty leaders from our emergency medicine (EM) residency, fellowship, and clerkship programs. Other non-leadership faculty, EM residents, and pediatric EM fellows with a focus on education also participated. The group was formed in 2008. Initially, the group was led by a successful basic scientist with a challenge to the educational faculty to look for scholarship in their day-to-day work. Skill building and faculty development (invited presentations on education research topics, participation of faculty in Medical Education Scholars Program were the focus of those first few years. She left to become division director elsewhere in 2009. The group's early work and focus on scholarship was propelled forward in late 2011 with the start of a new associate chair for education who had specific experience in education research and a track record of publishing educational scholarship. Thirteen faculty attend regularly, with about 8-10 present at each meeting. There is no mandate, incentive or tracking of participation.

The key components of each meeting include several areas. During the discussions of new projects, ideas are shared, research plans developed, and teams formed. Group mentoring occurs through the detailed discussions surrounding project development. The research work is implemented by the team outside of MERG, with updates and problems brought back to the group. These are often related to curriculum changes in the educational programs.

For each project there is an intentional process over sequential meetings. One of our projects was the outcome of a residency program change that entailed a switch from confidential to faculty-identified evaluation of residents. The MERG group decided to study the impact of this change, which resulted in the eventual completion of an abstract and manuscript submission. To achieve this kind of goal, we take these steps: 1) We develop the research questions, data to be collected, and determine who will compose the research team. 2) At each meeting we include an update of current projects to ensure continued project momentum and completion. 3) There is intentional scholarship planning surrounding national meetings. Several months before a national meeting submission deadline, we brainstorm and plan for didactic and research submissions including current and new projects. 4) If we identify a knowledge deficit or an educational need of the group, we will read an article or invite a local expert for the purpose of skill development. For example, a local expert on survey design was invited to present on key elements of successful survey research and also review ongoing projects.[6-8]

Clear Goals as a Coordinating Function through the Formation of a MERG

MERG meets monthly with the goal of bringing scholarly inquiry to the usual tasks required by medical educators such as curriculum design and trainee assessment, thereby turning our usual work into scholarship. This pushes the group to consider how the usual work of education can be scholarly. Thus, many changes and innovations in curriculum or educational processes are accompanied by hypothesis generation, data collection and analysis leading to research studies and other scholarship. This creates a distinctive culture of research emphasis and scholarly inquiry so that whenever we consider a change, there is the accompanying question of how we are going to measure the effectiveness. For example, when we began using Accreditation Council for Graduate Medical Education Milestones, several faculty queried how prepared graduating students would be for the new milestones and who was responsible for ensuring preparedness. We saw an opportunity to assess the preparedness and assessment of medical students during the transitions to internship. As a result, two projects were implemented and published.[9,10] Thus, we study the impact of educational innovations implemented in our programs and submit the work for peer-

review and dissemination.

Leadership with Expertise and Skill

Initially senior faculty with expertise in medical education as well as clinical or bench research expertise provided mentorship. As the group has gained experience, peer mentorship is predominant. As individuals develop areas of expertise, they are often tasked to present back to the group to share knowledge. In addition, we have begun to see the initial faculty participants in MERG taking the lead on their own independent projects and actively including junior members who are just making an initial foray into scholarship. The group now includes the educational leadership of all of the education domains from medical students, residents, and fellows, as well as individuals in other domains looking to cross into educational scholarship.

Team Science

We work as a team. Sharing projects has led to motivation, accountability, and work completion. Our monthly meetings serve as brainstorming sessions for new projects, research skill building, and tracking work completion. The team projects and monthly meetings serve to provide a positive culture with assertive participation and frequent communication as described by Bland. We intentionally include multiple author teams to create a division of labor, such as writing different sections of the manuscript or submitting the accompanying MedEDPortal publications, so that the workload, responsibility, and recognition are distributed among all team members.

Resources

Often faculty do not have the bandwidth or passion for writing required to bring manuscripts to completion, leading to failure of dissemination of educational scholarship. Therefore, when we realized that we had over 10 published abstracts that had not been turned into manuscripts we tried a new process to address this challenge. We theorized that we should be able to use a graduate student to help us translate our ideas and abstracts into published papers. The group worked with a Master's of Public Health graduate student for about 25 hours a month for eight months at low cost as it was subsidized by the federal work-study program. (The cost was less than $1,000 for about five hours a week). His role was to help bring the research ideas to completion through performance of the literature review, clarification of the study concept, data interpretation, and drafting of the manuscript. He completed the initial draft that was usually substantially revised by the first or anchor author who then coordinated the revisions and final product. He helped the group maintain a tight timeline to bring to completion one research project per month.

Recognizing that additional resources were needed to augment faculty effort, we undertook additional steps 1) We used undergraduate research assistants for data collection (three projects). As in many academic institutions, there are undergraduate students who can collect data. For example, one of the projects collected patient surveys scoring resident communication. If this resource is not available, then it would be important to steer away from projects that require specific hands-on data collection. 2) At times our residency administrative support was used for retrieving, organizing, or entering data. For example, an administrative assistant might download the faculty scoring of intern milestones for analysis. The amount of work required was variable. 3) It is often the last stages of submission that creates delays. For that purpose, we trained an administrative assistant to do referencing with EndNote, ™ maintain an education-based EndNote library, and perform final proofs of manuscripts. 4) Occasional unpaid statistical support was used from the Department of Medical Education (six projects). 5) As needed, we have used outside resources by inviting outside scholars to speak with the group. For example, we invited Dr. (name blinded) to assist us with standard setting our competency examination and Dr. (name blinded) to clarify the process of MedEdPortal submissions. These resources were accelerators to publication without which the group would have been much slower to publish.

IMPACT & EFFECTIVENESS

Impact

To measure the impact of this educational innovation, we collected the following information for the MERG group as a whole and for each faculty member: number of abstracts accepted at regional, national, or international meetings, didactic presentations, papers accepted or published, papers submitted but under review or not yet accepted, grants, and promotions from 2011-2014. We excluded book chapters because they are not peer reviewed. Studying the impact of MERG was determined not to be regulated by the IRB.

Over the past four years, the MERG group has been highly productive. The team effort resulted in 102 presented abstracts, 46 publications, and 35 didactic presentations (Table). We anticipate additional publications as a number of papers are currently under review. In addition, we have encouraged trainee scholarship. Indirect impact through skills development and scholarship was also evident in the fact that members of the group are contributing to four grants totaling over 10 million dollars (grant sources: Simulation Center-internal, Interprofessional Center-internal, the Department of Defense, the American Medical Association).

The intentional focus on medical education research has led our collaborative group of educators to successfully promote our scholarship, which will contribute to faculty promotion. (Five members of our group have been promoted.) In addition, because we are deliberate in our assessment of all our educational innovations we are able to refine our curricula and ultimately create a better learning environment for our trainees. The model, following Bland's research guideline of

Table. Summary of impact on academic output after the establishment of a collaborative group of emergency medicine educators.

Scholarship for the 13 MERG faculty	Average per MERG member* (SD)	Total for the group
Abstract presentations (+/-publication)	14 (17)	102
Publications (accepted or published)	4.9 (5.7)	46
Papers in review or revision	3.4 (2.9)	14
MedEd portal publications	2.2 (2.9)	13
MedEdPortal in review		2
Scholarship with students, residents, fellows		60
Didactic sessions	3.5 (4.9)	35
Total scholarship (limitation: this number may double count abstracts and publications)	28.7 (32.1)	212
Faculty members promoted (4 instructor to assistant, 1 associate to full)		5
Grants		4

MERG, medical education research group; *SD,* standard deviation
*Average is calculated by total number divided by the number of MERG members (13).

creating a group with clear goals that meets monthly to work as a team and adds resources as needed, was found to be both feasible and effective.

Three key attributes of this model lie at the heart of its success. First, this group facilitates valuable mentorship between members that may not have taken place otherwise. Mentorship, both by senior faculty and by peers, is a vital aspect of growth and learning and has allowed group members to develop into more highly-skilled researchers.[11-15] Second, this group facilitates effective teamwork. This teamwork allows new faculty to get involved in research projects more easily, keeps faculty accountable to each other for producing results, and provides a venue for creating, quickly vetting and refining research ideas. Third, this group identifies education gaps within the group and addresses these needs through presentations from visiting scholars, discussions within the group, reference materials and articles. This creates a community of educational practice and level of discussion because members are educated on topics they would not have had the opportunity to learn about otherwise and raises the level of discussion and implementation.

It is our hope that this innovation will inspire other institutions to create new education research groups based on this model. To that aim, a number of key challenges that this group has faced are outlined below. These lessons can provide insights for other institutions into how to create research groups of their own.

A major benefit of this group, as outlined above, is that it facilitates mentoring between members. Due to time constraints on medical students and residents it can be difficult for them to participate. Increased effort needs to be employed to include trainees in projects and meetings. Another challenge was that due to the small and resource-limited nature of this group, demand for administrative support was at times higher than available capacity.

A further challenge is that such a group may require an organizational catalyst or educational expert to drive formation and commitment until the cultural change is established. This leadership may be available through virtual mentorship or use of non-education research experience or educationalists outside the department. Additional difficulties may be faced by smaller and more resource-limited institutions wanting to establish a similar research model. However, we found that this model did not consume significant resources and was effective at promoting scholarly activity. We believe, therefore, that this innovation presents a useful method of increasing academic output in any emergency department that wishes to implement it.

CONCLUSION

An intentional focus on scholarship has led to our collaborative group of educators successfully increasing their scholarship through team productivity, which ultimately leads to faculty promotions.

Address for Correspondence: Sally Santen, MD, PhD, University of Michigan, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-5305. Email: ssanten@umich.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Love JN, Coates WC, Santen SA, et al. The MERC at CORD Scholars Program in Medical Education Research: A Novel Faculty Development Opportunity for Emergency Physicians. *Acad Emerg Med.* 2009;16(S2):S37-S41.
2. Kessler C and Burton JH. Moving Beyond Confidence and Competence: Educational Outcomes Research in Emergency Medicine. *Acad Emerg Med.* 2011;18:S25-S26.
3. Yarris LM and Deiorio NM. Education Research: A Primer for Educators in Emergency Medicine. *Acad Emerg Med.* 2011;18:S27-S35.
4. Bland CJ, Hitchcock MA, Anderson WA, et al. Faculty development fellowship programs in family medicine. *J Med Educ.* 1987;62(8):632-41.
5. Bland CJ and Ruffin MT. Characteristics of a productive research environment: literature review. *Acad Med.* 1992;67(6):385-97.
6. Moore DE, Green JS, Gallis HA. Achieving desired results and improved outcomes: Integrating planning and assessment throughout learning activities. *J Contin Educ Health.* 2009;29(1):1-15.
7. van der Vleuten C. Validity of final examinations in undergraduate medical training. *Brit Med J.* 2000;321(7270):1217-19.
8. Andreatta PB and Gruppen LD. Conceptualising and classifying validity evidence for simulation. *Med Educ.* 2009;43(11):1028-35.
9. Santen SA, Peterson WJ, Khandelwal S, et al. Medical student milestones in emergency medicine. *Acad Emerg Med.* 2014;21(8):905-11.
10. Santen SA, Rademacher N, Heron SL, et al. How competent are emergency medicine interns for Level 1 milestones: Who is responsible? *Acad Emerg Med.* 2013;20(7):736-9.
11. Anderson KD and Mavis BE. The relationship between career satisfaction and fellowship training in academic surgeons. *Amer J Surgery.* 1995;169(3):329-33.
12. Broaddus VC and Feigal Jr DW. Starting an academic career. a survey of junior academic pulmonary physicians. *Chest.* 1994;105(6):1858-63.
13. Gitlin SD, Yuan Z, Little RJ, et al. Factors that influence successful training and faculty career development in hematology/oncology patient-oriented clinical research. *J Cancer Educ.* 2005;20(2):72-8.
14. Neacy K, Stern SA, Kim HM, et al. Resident perception of academic skills training and impact on academic career choice. *Acad Emerg Med.* 2000;7(12):1408-15.
15. Taylor JS, Friedman RH, Speckman JL, et al. Fellowship training and career outcomes for primary care physician-faculty. *Acad Med.* 2001;76(4):366-72.

Implementation of an Education Value Unit (EVU) System to Recognize Faculty Contributions

Joseph House, MD*
Sally A. Santen, MD, PhD†
Michele Carney, MD*
Michele Nypaver, MD*
Jonathan P. Fischer, MPH‡
Laura R. Hopson, MD*§

*University of Michigan Medical School, Department of Emergency Medicine Children's Emergency Services, Ann Arbor, Michigan
†University of Michigan Medical School, Department of Emergency Medicine, Department of Learning Health Sciences, Ann Arbor, Michigan
‡University of Michigan, School of Public Health, Ann Arbor, Michigan
§University of Michigan Medical School, Department of Emergency Medicine, Ann Arbor, Michigan

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted April 1, 2015; Revision received August 3, 2015; Accepted August 18, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.8.26136

Introduction: Faculty educational contributions are hard to quantify, but in an era of limited resources it is essential to link funding with effort. The purpose of this study was to determine the feasibility of an educational value unit (EVU) system in an academic emergency department and to examine its effect on faculty behavior, particularly on conference attendance and completion of trainee evaluations.

Methods: A taskforce representing education, research, and clinical missions was convened to develop a method of incentivizing productivity for an academic emergency medicine faculty. Domains of educational contributions were defined and assigned a value based on time expended. A 30-hour EVU threshold for achievement was aligned with departmental goals. Targets included educational presentations, completion of trainee evaluations and attendance at didactic conferences. We analyzed comparisons of performance during the year preceding and after implementation.

Results: Faculty (N=50) attended significantly more didactic conferences (22.7 hours v. 34.5 hours, $p<0.005$) and completed more trainee evaluations (5.9 v. 8.8 months, $p<0.005$). During the pre-implementation year, 84% (42/50) met the 30-hour threshold with 94% (47/50) meeting post-implementation ($p=0.11$). Mean total EVUs increased significantly (94.4 hours v. 109.8 hours, $p=0.04$) resulting from increased conference attendance and evaluation completion without a change in other categories.

Conclusion: In a busy academic department there are many work allocation pressures. An EVU system integrated with an incentive structure to recognize faculty contributions increases the importance of educational responsibilities. We propose an EVU model that could be implemented and adjusted for differing departmental priorities at other academic departments. [West J Emerg Med. 2015;16(6):952–956.]

INTRODUCTION

Changes in healthcare have placed pressure on emergency departments (EDs). For academic EDs, this presents added challenges as they struggle to balance their tripartite missions of clinical care, research, and education. There are often fewer incentives for educational activities than for other domains, and as a result, education may be given a lower priority.[5]

Efforts have been made to better align departmental budgets between clinical care, research, and education.

"Mission-based budgeting" began in 1999 and has grown in popularity.[1,2] While this system allocated more resources to educational activity, departments struggled to equitably distribute these funds to individual faculty. The result was often that incentives were not tied to specific education-related activity, but were evenly distributed among faculty.[2] In response, medical schools attempted to quantify educational activity using the relative value unit system of measuring patient care activity as a model.[2-5] The educational value unit (EVU), although promising in its potential to incentivize educational activity, has not achieved widespread utilization or been studied extensively. This is especially true in the ED, where only one study was published a decade ago.[5]

In 2011, our chair established a Faculty Incentive Task Force that included faculty representing all departmental missions within our academic ED. Faculty contribution to the educational mission was identified as a core metric, and an educational subcommittee was formed to review the available literature and other departmental practices to develop measurement criteria. The purpose of this study was to determine the feasibility of an EVU system in an academic ED and to examine its effect on faculty behavior in the educational mission.

METHODS

Study Design and Setting

This is a prospective observational study that was reviewed by the IRB and determined to be exempt.

Methods and Measurements

Through group consensus, the Faculty Incentive Task Force agreed upon broad priorities supporting education. These included providing lectures, conference participation, participation in trainee recruitment, and completion of trainee assessments. An analysis of all educational activity performed during the prior academic year was performed, which allowed identification of common activities as well as creation of a tracking process.

Four main educational activity categories for EVUs (measured in hours) were created for educational contribution to the department. Activities were informed by medical school departmental funding models, and each activity was assigned a standardized time value determined by group consensus (Table 1). In general, preparing and leading an educational session de novo, earned more hours than presenting an existing lecture or assisting in a conference. Value was also given to activities requiring faculty time, including trainee recruitment as well as conference attendance. The "additional" teaching category included educational activities such as teaching in other departments, mentoring, and involvement in educational committees. It relied on faculty self-report. This final category was incorporated during the post-implementation year based on faculty feedback and was excluded from the analysis.

We used administrative data collection to track the four intradepartmental categories. These were tracked from existing materials from the residency, fellowship and student domains, including final conference schedules, attendance records, and Medhub reports. We relied on faculty input to complete the final "additional" teaching category. In addition, we created mechanisms to regularly update faculty as to their participation levels. The first iteration of the EVU calculation was tested against faculty activity from the immediate past academic year, and an initial benchmark selected which would allow the majority of faculty to meet based on existing activities in order to foster support for the program during implementation. The program was implemented in July 2013.

Outcomes and Analysis

After implementation of the program, we compared impact on educational priority items (e.g. conference attendance, completion of resident evaluations) to pre-implementation levels using paired t-tests. Data analysis was performed using Microsoft Excel 2010 and graphpad.com calculators. We analyzed EVU achievement in total and all individual domains . Completion rates for resident and fellow evaluations by faculty for July 2012-May 2013 (11 pre-implementation months) and July 2013-May 2014 (11 post-Implementation months) were compared. We defined a faculty member as completing evaluations for the month if they completed at least one resident or fellow evaluation during a given month. Faculty were responsible for determining which trainees they were able to evaluate.

RESULTS

We included in the analysis (N=50) faculty members from the children's and adult divisions of the ED who were working clinically during both the pre- and post-implementation periods and subject to the incentive program

The differences between the pre-implementation period and the EVU measurement period are presented in Table 2. Faculty attended significantly more didactic conferences and completed more resident monthly evaluations. The majority of faculty members (84%) met the 30-hour threshold for compliance in the pre-implementation year and 94% in the post-implementation year. There was a small but significant increase in total EVUs between the years due to increased conference attendance and evaluation completion. Interestingly, faculty members who tend not to complete resident evaluations did not change with the new system; of the eight faculty members who completed zero resident evaluations in either academic year, five had zero both years. In contrast, faculty members who completed some evaluations tended to do more under the incentive system. There were no differences between the periods in EVUs attained for educational presentations and trainee recruitment.

Overall, administrative time required was estimated at one hour per month in each of the three major domains

Table 1. Education value unit per educational activity.

Departmental activities-baseline	Expected minimums
1) Leading a new educational sessions (including residency or fellowship lectures, administrative track or other departmental seminar, or intern orientation lecture(10 hours of time to account for preparation for each 1 hour presented) Or Preparing & leading an active learning session: (e.g. Skills & procedure labs, simulation, oral board exams/ clinical skills exams cases, small group sessions, focused residency retreats, Peds OSCEs (5 hours credit:1 hour presented) Or Assisting with active learning sessions, small groups, skills labs, mentoring resident session, focused mentoring (2 hours credit:1 hour presented) Or Student teaching sessions (1 hour credit:1 hour presented) Or Other teaching activities EMIG, clinical skills assessments (CSA) (1 hour:1 hour)	10 hours
2) Didactic conference attendance (1:1) optimal 3hrs/month=36	10 hours
3) Completion of evaluations of residents & fellows (1 hour per month)	10 hours
4) Recruitment interviews for residency or fellowship programs (1 hour:1 hour)	
5) Additional activities-resident mentoring, educational activities outside department, educational committees (maximum 10 hrs.)	
Total (minimum expected)	Total=30 hrs

Peds, pediatrics; *OSCEs*, objective structured clinical examination; *EMIG*, emergency medicine interest group

(student, residency, and fellowships) and approximately 10 hours annually for development and distribution of biannual progress reports to faculty and departmental leadership.

DISCUSSION

This study demonstrated that a system of applying value to educational activity is feasible. Additionally, data indicate a positive effect on physician participation while providing objective data for reward for effort supporting educational activities. It is worthwhile to note that despite the small increase in overall educational activities (EVUs), those increases were confined to the smaller scale activities such as evaluation completion and conference attendance. One may speculate that the larger time investment involved in creating a conference presentation is not felt worth the investment despite the weighting of hours in the EVU system.

A characteristic of this system that likely contributed to its success is that it relies on multiple sources of motivation to change faculty behavior. The most obvious motivation was a financial incentive. Although it is commonly believed that financial incentives can change behavior, the literature studying the effect of financial incentives on primary care physician behavior has been inconclusive.[6] In the current study, the maximum financial bonus related to educational activities was approximately 1% of annual salary. The EVU represented one component of an eight-part overall ED incentive system representing educational, scholarly, and clinical goals each with an equal weight applied universally to all faculty.

While faculty response to the amount of the bonus was beyond the scope of this study, it is not clear that the financial incentive itself was sufficient to explain all of the results. An additional motivation may also be the desire for high achievement or a Hawthorne-type effect. This system made educational activity an explicit priority. It also included EVUs as part of each physician's annual review with the chair. As such, achieving high EVUs was more closely associated with desired performance.

The final motivating factor was likely peer encouragement. Participation in some of the educational activities was made public with monthly summaries. As such, there may have been pressure for physicians to attend events to increase the publicly displayed attendance.

The combined effects of these three sources of motivation are what likely brought about our positive results. Interestingly, our results were somewhat mixed as metrics in prioritized subcategories improved, but overall EVUs achieved experienced only a slight increase. This suggests that individuals tended to put their emphasis on the categories that were most easily recognized. Further, there was no increased incentive to go beyond the threshold EVU, although most faculty did. However, it should be noted that participation in all domains was stable or increased, so we were able to cumulatively increase faculty participation.

This study builds on the work of Khan and Simon, who developed a system in 2003 to quantify and reward teaching activity in an ED.[5] Their system applied weights to activities that were deemed more essential, as opposed to our system that was based on the time required to complete each task. While the Khan system included only teaching activities for medical students, our system included a wide array of

Table 2. Year-to-year comparison of education value unit (EVU) results pre and post implementation. Total EVUs increase as a result of increased activity with the conference attendance and evaluation completion subcategories. Participation in recruitment activities and educational sessions are unchanged. The final category of non-departmental educational activities was in effect only for the post-implementation year so was excluded from the analysis. Domains involved in each subcategory include residency, fellowships, and medical students.

	Total EVUs (hours) attained	EVUs (hours) earned from all departmental educational sessions (R,F,S)	EVUs (hours) earned from didactic conference attendance (R,F)	EVUs (hours) from completion of evaluations (R, F)	EVUs (hours) from educational recruitment activities (R,F)
Pre-implementation year	Mean 94.4 SD 75 Range 0-390	Mean 56.9 SD 52.2 Range 0-239	Mean 22.7 SD 16.8 Range 0-75	Mean 5.9 SD 3.8 Range 0-16	Mean 8.8 SD 19.4 Range 0-12
Post-implementation year	Mean 109.8 SD 90 Range 0-521	Mean 58.5 SD 68.7 Range 0-369	Mean 34.5 SD 19.9 Range 0-109	Mean 8.8 SD 6.4 Range 0-26	Mean 7.9 SD 16.3 Range 0-96
95% CI	[-17.8, 48.3]	[-22.6, 25.8]	[4.5, 19.1]*	[1.1, 2.0]*	[-8.0, 6.2]
ρ-value	0.04*	n.s.	<0.005*	<0.005*	n.s.

EVU, education value unit; *R*, residency; *F*, fellowships; *S*, medical students
*significant p value

educational activities to recognize diverse contributions. Although pre-implementation data were not available, Khan and Simon report progressive increases in both group and individual productivity during the first three years of implementation. Our results build on previous literature suggesting that incentive systems may be effective at increasing educational activity in an academic ED.

LIMITATIONS

There are several limitations to this study. First, it is a single site focused on initial response to the implementation of the EVU system and generalizability may be limited. It would be difficult to implement identical EVU systems across multiple EDs; thus, we encourage others to customize EVU systems and study the results. Second, it is difficult to determine the exact motivations of faculty behaviors. Whether the EVU set the tone for a culture recognizing the importance of education or whether there were other factors at play is not known. Third, we did need to balance administrative feasibility with faculty desire to recognize all possible educational activities. Finally, organizational politics had an effect on the EVU program with the decision to emphasize departmental activities rather than extra-departmental.

CONCLUSION

As external time and financial pressures continue to increase, it is imperative that academic EDs remain committed to their educational missions. Achieving this will require innovative methods to use limited resources. Once developed, the EVU system may be tailored to address changing departmental priorities and challenges. This study demonstrates that development of an EVU system to incentivize faculty is feasible and effective in motivating faculty to meet educational responsibilities. This study represents an important step in that direction and hopefully will prompt further investigation into how to best promote educational activity in busy academic EDs.

Address for Correspondence: Laura R. Hopson, MD, University of Michigan Medical School, 1500 E. Medical Center Drive, TC B1 380-0305, Ann Arbor, MI 48109-5305. Email: Lhopson@med.umich.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Watson RT and Romrell LJ. Mission-based budgeting: removing a graveyard. *Acad Med.* 1999;74:627-40.
2. Stites S, Vansaghi L, Pingleton S, et al. Aligning compensation with education: design and implementation of the Educational Value Unit (EVU) system in an academic internal medicine department *Acad Med*. 2005;80:1100-6.
3. Williams RG, Dunnington GL, Folse JR. The impact of a program for systematically recognizing and rewarding academic performance. *Acad Med*. 2003;78:156-66.
4. Clyburn EB, Wood C, Moran W, et al. Valuing the education mission: implementing an educational value units system. *Am J Med.* 2011;124:567-72.

5. Khan NS and Simon HK. Development and implementation of a relative value scale for teaching in emergency medicine: the teaching value unit. *Acad Emerg Med.* 2003;10:904-7.

6. Scott A, Sivey P, Ait Ouakrim D, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev.* 2011:CD008451.

Correlation of the National Board of Medical Examiners Emergency Medicine Advanced Clinical Examination Given in July to Intern American Board of Emergency Medicine in-training Examination Scores: A Predictor of Performance?

Katherine Hiller, MD, MPH*
Doug Franzen, MD, MEd†
Corey Heitz, MD‡
Matthew Emery, MD§
Stacy Poznanski, DO¶

* University of Arizona, Department of Emergency Medicine, Tucson, Arizona
† University of Washington, Department of Emergency Medicine, Seattle, Washington
‡ Virginia Tech Carilion, Department of Emergency Medicine, Roanoke, Virginia
§ Michigan State University College of Human Medicine, Department of Emergency Medicine, East Lansing, Michigan
¶ Wright State University, Department of Emergency Medicine, Dayton, Ohio

Section Editor: Lalena M. Yarris, MD
Submission history: Submitted May 1, 2015; Revision received September 4, 2015; Accepted September 9, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.9.27303

Introduction: There is great variation in the knowledge base of Emergency Medicine (EM) interns in July. The first objective knowledge assessment during residency does not occur until eight months later, in February, when the American Board of EM (ABEM) administers the in-training examination (ITE). In 2013, the National Board of Medical Examiners (NBME) released the EM Advanced Clinical Examination (EM-ACE), an assessment intended for fourth-year medical students. Administration of the EM-ACE to interns at the start of residency may provide an earlier opportunity to assess the new EM residents' knowledge base. The primary objective of this study was to determine the correlation of the NBME EM-ACE, given early in residency, with the EM ITE. Secondary objectives included determination of the correlation of the United States Medical Licensing Examination (USMLE) Step 1 or 2 scores with early intern EM-ACE and ITE scores and the effect, if any, of clinical EM experience on examination correlation.

Methods: This was a multi-institutional, observational study. Entering EM interns at six residencies took the EM-ACE in July 2013 and the ABEM ITE in February 2014. We collected scores for the EM-ACE and ITE, age, gender, weeks of clinical EM experience in residency prior to the ITE, and USMLE Step 1 and 2 scores. Pearson's correlation and linear regression were performed.

Results: Sixty-two interns took the EM-ACE and the ITE. The Pearson's correlation coefficient between the ITE and the EM-ACE was 0.62. R-squared was 0.5 (adjusted 0.4). The coefficient of determination was 0.41 (95% CI [0.3-0.8]). For every increase of one in the scaled EM-ACE score, we observed a 0.4% increase in the EM in-training score. In a linear regression model using all available variables (EM-ACE, gender, age, clinical exposure to EM, and USMLE Step 1 and Step 2 scores), only the EM-ACE score was significantly associated with the ITE ($p<0.05$). We observed significant colinearity among the EM-ACE, ITE and USMLE scores. Gender, age and number of weeks of EM prior to the ITE had no effect on the relationship between EM-ACE and the ITE.

Conclusion Given early during intern year, the EM-ACE score showed positive correlation with ITE. Clinical EM experience prior to the in-training exam did not affect the correlation. [West J Emerg Med. 2015;16(6):957–960.]

INTRODUCTION

Incoming interns to emergency medicine (EM) residencies come from a variety of educational backgrounds, creating significant variations in their baseline funds of knowledge. Program directors must quickly ascertain if any interns have unusual knowledge gaps or learning difficulties that may require a specialized learning plan or remediation. Traditionally, the first high quality, objective testing available for assessment of interns has been the American Board of EM (ABEM) in-training examination (ITE), which is offered annually on the last Wednesday in February. According to the ABEM website, "It is a standardized examination that residents and program faculty can use to judge an individual resident's progress toward successful ABEM certification. There is a strong relationship between in-training and qualifying examination scores. Physicians with higher in-training scores have a higher likelihood of passing the qualifying examination and those with lower scores have a lower likelihood of passing the qualifying examination."[1] This statement is supported by an observed moderate correlation between post-graduate year (PGY) 3 ITE scores and ABEM written examination scores.[2] In addition to providing predictive information to program directors regarding the residents they are about to graduate, the ITE also provides national norms for residents at all other PGY levels of training.

United States Medical Licensing Examination (USMLE) scores also provide information for program directors about their incoming interns' baseline knowledge. Step 1 scores are mildly correlated (R2 0.25) and Step 2 scores are moderately correlated (R2 0.43) with the EM ITE.[3] However, USMLE Step 2 exams are typically taken in the fall of the fourth year of medical school (M4) and the ITE is not given until late winter. This gap of roughly 18 months is also one of the most variable periods in medical education in both content and clinical exposure for graduating M4s/incoming interns.[4] Some program directors have responded to the disparity in incoming interns by providing "boot camps" to their interns immediately before, during or after orientation in July.[5-7]

In April 2013, the National Board of Medical Examiners (NBME) released the EM Advanced Clinical Examination (EM-ACE).[8] This is an examination based on the national fourth year EM medical student curriculum first published in 2006 and updated in 2010.[9,10] It is intended to be administered to fourth year students at the end of their EM rotation. To rapidly administer the examination and calculate scaled scores and internal validity, the NBME offered the EM-ACE free of charge for the first year of administration. In response, several residency program directors administered the EM-ACE to their incoming interns in July to identify interns who required additional educational exposure or attention. However, there is no data on whether performance on the EM-ACE, when administered to EM residents at the onset of internship, has any predictive value to known outcome measures such as ITE performance.

The objective of this study was to determine the correlation, if any, between intern scores on the EM-ACE administered in July, and intern scores on the ITE administered the subsequent February. In addition, we sought to assess whether USMLE scores correlate with EM-ACE scores when administered at the onset of internship, as USMLE scores have been shown to correlate with intern ITE scores.

METHODS

This study was approved by the institutional review board of each participating residency, and was determined exempt from human subjects review.

This was a multi-institutional, observational study. In July 2013, entering EM interns at six geographically diverse residency programs took the EM-ACE. They underwent standard residency training, and in February 2014 took the required ABEM ITE. Scaled examination scores for both the EM-ACE and the internship ABEM ITE score were collected electronically from program coordinators and/or program directors. Additional data collected included the date of EM-ACE administration, institution, gender, age, USMLE Step 1 and Step 2 scores, and number of weeks of clinical EM and off-service experience completed during their current residency program prior to taking the ITE.

We performed linear regression to determine the relationship between EM-ACE scores and the EM ITE scores, and calculated Pearson's correlation coefficients. Data was collected with Microsoft Excel 2007 and analyzed using StataMP-11 (College Station, TX).

RESULTS

A total of 62 interns took the ITE at six residency programs. Of these, 60 (96.8%) also took the EM-ACE in July of their intern year. Two residents were sick on the date of EM-ACE administration, both from the same institution. Data were available for USMLE Step 1 in 50 (80.6%) and Step 2 in 48 (77.4%) of these residents. Scores on the Comprehensive Osteopathic Medical Licensing Examination (COMLEX) 2 were available for six (9.7%) of the residents. See Table 1.

Gender was slightly skewed towards male (58.06%), and the average age of the interns at examination administration was 30 years old. On average, interns had experienced 17 weeks (SD 4.4) of clinical time in the emergency department (ED) from the start of EM residency to the ITE. The average EM-ACE score was 69.8 (SD 7.1), and the average ITE score (percent correct) was 70.5% (SD 8.4%).

In a linear regression model using all the available variables, gender (p=0.99), age (p=0.52) and clinical exposure to EM prior to the ITE (p=0.53) were not associated with the in-training score. USMLE Step 1 (p=0.61) and Step 2 (p=0.53) were likewise not associated with ITE score. There were too few COMLEX scores to allow incorporation into the linear regression model.

The EM-ACE score was significantly associated with the ITE score ($p<0.05$). The coefficient of determination was 0.41 (95% confidence interval 0.3-0.8); in other words, for every

Table 1. Demographics of interns completing the emergency medicine advanced clinical examination (EM-ACE) and the in-training exam.

Demographics	Value
Gender, male	58.1%
Age, years	30.4 (4.3)
USMLE, step 1 (n=50)	225 (18.0)
USMLE, step 2 (n=48)	237.7 (17.8)
Weeks of EM prior to in-training	17.6 (4.4)
EM-ACE (scaled) (n=60)	69.8 (7.1)
EM in-training (n=62)	70.5% (8.4%)

USMLE, united states medical licensing examination

increase of one in the scaled EM-ACE, we observed a 0.4% increase in the EM in-training score. The Pearson's correlation coefficient between the EM ITE score and the EM-ACE was 0.62. R-squared value was 0.5 (adjusted 0.4).

In a regression model only containing the EM-ACE (the only variable significant in the full model), the significance was even higher ($p<0.001$, coefficient of determination 0.7 (0.4-1.0); however, the correlation was lower (R-squared 0.38). Interestingly, a similar pattern of independent association with the ITE was observed with USMLE Step 1 scores ($p<0.001$, R-squared 0.39) Step 2 (<0.001, R-squared 0.33) and COMLEX scores ($P<0.05$, R-squared 0.73) when used alone in a linear regression model predicting ITE score. In a model with all three exams (EM-ACE, Step 1 and Step 2 [there were too few COMLEX scores to include in the limited regression model]), the EM-ACE ($p<0.05$) and Step 1 ($p<0.05$) were still significantly associated with the in-training score, but Step 2 scores were not ($p=0.43$)

There was significant colinearity observed among the EM-ACE, the EM ITE and both USMLE Step scores (Table 2).

DISCUSSION

To provide effective education to learners, it is important to first assess their baseline knowledge. The general nature of the medical knowledge assessed by the USMLE, as well as the variation in timing of administration, especially as related to the timing of a student's EM rotation, makes the USMLE a less specific assessment of basic EM knowledge. Additionally, clinical experience after taking the USMLE is highly variable, ranging from one EM rotation in medical school to post-graduate experience.[4] As a result, incoming interns may have widely different clinical EM exposure and expertise. The ABEM ITE is an excellent tool for assessment of resident knowledge, and is predictive of performance on the EM qualifying examination. Program directors have been using the ITE as a means to assess their learners' progress towards competency since 1985.[11] However, the date of administration is fixed, and is set eight months into a 36-48 month residency.

The EM-ACE administered at the onset of residency, halfway through the most variable 18 months of EM training, correlates well with internship ITE scores, and may provide an earlier assessment of knowledge than the in-training exam. Identifying below-average performers is of particular importance to program directors, as early identification of these learners makes early intervention with a specialized learning plan possible and allows more time for remediation. In addition to assisting program directors with identification of potential problem learners, once baseline performance on the EM-ACE is known, the ITE could then serve as an assessment of teaching methods.

Interestingly, despite the observed variation in clinical experience at the six residencies, clinical ED experience did not affect the correlation between EM-ACE and ABEM examination performance. There are a number of potential explanations for this finding. First, and most obvious, as neither the EM-ACE nor the ABEM ITE assess clinical competency, it may be that clinical experience and exposure have a much greater effect on assessment methods that are sensitive to gains in clinical competence. It may also be that programs with less clinical exposure augmented their residents' learning by non-EM clinical activities (simulation, off-service rotations, didactics, self-learning) and vice versa. It is also possible that programs which assigned residents to less early EM-based clinical time had a greater focus on efficient learning in the limited time residents had in the ED.

LIMITATIONS

There are a number of limitations to this study. The EM-ACE was not intended for use as a "pre-test" for internship. It was intended as a high stakes examination for fourth-year EM medical students, and is based on the national fourth-year medical student curriculum.[9] The ITE is based on the EM model curriculum[12]_ENREF_11. While there is a large amount of overlap between these two curricula, they are not identical.

Performance on the EM-ACE (when used at the beginning of residency) is potentially affected by the weight an intern perceives a program director places on it. A lack of preparation could have negatively affected performance on the EM-ACE, as compared with a relatively augmented score on the ITE if

Table 2. Pearson's correlation coefficients between the United States medical licensing examination (USMLE) Step 1 and Step 2, the emergency medicine advanced clinical exam (EM-ACE) given in July of EM internship and the internship EM in-training exam score.

Exam	USMLE Step 1	USMLE Step 2	EM-ACE	In-training exam
USMLE step 1	N/A	0.70	0.60	0.62
USMLE step 2	0.70	N/A	0.64	0.58
EM-ACE	0.60	0.64	N/A	0.62
In-training exam	0.62	0.58	0.62	N/A

an intern prepared more for the latter. Some of the incoming interns may have already taken one form of EM-ACE. This is unlikely, as the EM-ACE was offered from April 2013 on; however, if a student took an EM rotation at the end of their fourth year, there is a possibility that student could have taken it twice. Limitations in the NBME scaling process itself may affect the EM-ACE scores observed. In the initial offering of the exam, a "reference group" of fourth-year medical students from LCME-accredited medical schools who were taking the EM-ACE for the first time was used to scale the exam. By definition, incoming interns differ from the reference population at least by level of training, and may differ from the reference group even further if they had taken the EM-ACE before, or attended a non-LCME-accredited medical school. Finally, there was significant colinearity observed among Step 1, Step 2, the EM-ACE and the ABEM ITE. This is likely due to the fact that students who perform well (or poorly) on standardized testing will perform well (or poorly) on all standardized tests. Standardized tests, and student/resident performance on them, do not necessarily assess competency. However, as board certification in our specialty hinges on ABEM qualifying exam performance, standardized examination performance is a proxy measure that holds value.

CONCLUSION

Performance on the EM-ACE given at the onset of residency correlates well with intern ABEM ITE performance. Earlier assessment of residents' fund of knowledge may provide program directors with an opportunity for earlier identification of residents with knowledge gaps and increased time to formulate specialized learning plans.

Address for Correspondence: Katherine M. Hiller, MD, MPH, University of Arizona, 1501 N. Campbell Ave, Tucson, AZ 85724. Email: khiller@aemrc.arizona.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. ABEM. American Board of Emergency Medicine In-training Examination Overview. Available at: https://www.abem.org/public/emergency-medicine-training/in-training-examination/in-training-examination-overview. Accessed May 1, 2015.
2. Frederick RC, Hafner JW, Schaefer TJ, et al. Outcome measures for emergency medicine residency graduates: do measures of academic and clinical performance during residency training correlate with American Board of Emergency Medicine test performance? *Acad Emerg Med.* 2011;18Suppl2:S59-64.
3. Thundiyil JG, Modica RF, Silvestri S, et al. Do United States Medical Licensing Examination (USMLE) scores predict in-training test performance for emergency medicine residents? *J Emerg Med*. 2010;38(1):65-9.
4. Barzansky B and Etzel SI. Medical schools in the United States, 2008-2009. *JAMA*. 2009;302(12):1349-55.
5. Min AA, Stoneking LR, Grall KH, et al. Implementation of the Introductory Clinician Development Series: an optional boot camp for Emergency Medicine interns. *Adv Med Educ Pract*. 2014;5:275-9.
6. Krajewski A, Filippa D, Staff I, et al. Implementation of an intern boot camp curriculum to address clinical competencies under the new Accreditation Council for Graduate Medical Education supervision requirements and duty hour restrictions. *JAMA Surg*. 2013;148(8):727-32.
7. Fernandez GL, Page DW, Coe NP, et al. Boot cAMP: educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. *J Surg Educ*. 2012;69(2):242-8.
8. NBME. 2014-2015 Subject Examination Fees. 2014. Accessed June 7, 2014.
9. Manthey DE, Ander DS, Gordon DC, et al. Emergency medicine clerkship curriculum: an update and revision. *Acad Emerg Med*. 2010;17(6):638-43.
10. Manthey DE, Coates WC, Ander DS, et al. Report of the Task Force on National Fourth Year Medical Student Emergency Medicine Curriculum Guide. *Ann Emerg Med*. 2006;47(3):e1-7.
11. Gillen JP. Structured emergency medicine board review and resident in-service examination scores. *Acad Emerg Med*. 1997;4(7):715-17.
12. 2013 Model of the Clinical Practice of Emergency Medicine. 2013; Available at: https://www.abem.org/public/docs/default-source/publication-documents/2013-em-model---website-document-pdf.pdf?sfvrsn=8. Accessed May 1, 2015.

Effect of a Novel Engagement Strategy Using Twitter on Test Performance

Amanda L. Webb, MS*
Adam Dugan, MS†
Woodrow Burchett, MS†
Kelly Barnett, MD‡
Nishi Patel, MD‡
Scott Morehead, MD‡
Mark Silverberg, MD‡
Christopher Doty, MD‡
Brian Adkins, MD‡
Lauren Falvo, BA*

*University of Kentucky College of Medicine, Lexington, Kentucky
†University of Kentucky, Department of Statistics, Lexington, Kentucky
‡University of Kentucky, Department of Emergency Medicine, Lexington, Kentucky

Section Editor: David Wald, DO
Submission history: Submitted April 30, 2015; Revision received October 19, 2015; Accepted October 26, 2015
Electronically published November 12, 2015
Full text available through open access at http://escholarship.org/uc/uciem_westjem
DOI: 10.5811/westjem.2015.10.28869

Introduction: Medical educators in recent years have been using social media for more penetrance to technologically-savvy learners. The utility of using Twitter for curriculum content delivery has not been studied. We sought to determine if participation in a social media-based educational supplement would improve student performance on a test of clinical images at the end of the semester.

Methods: 116 second-year medical students were enrolled in a lecture-based clinical medicine course, in which images of common clinical exam findings were presented. An additional, optional assessment was performed on Twitter. Each week, a clinical presentation and physical exam image (not covered in course lectures) were distributed via Twitter, and students were invited to guess the exam finding or diagnosis. After the completion of the course, students were asked to participate in a slideshow "quiz" with 24 clinical images, half from lecture and half from Twitter.

Results: We conducted a one-way analysis of variance to determine the effect Twitter participation had on total, Twitter-only, and lecture-only scores. Twitter participation data was collected from the end-of-course survey and was defined as submitting answers to the Twitter-only questions "all or most of the time", "about half of the time", and "little or none of the time." We found a significant difference in overall scores ($p<0.001$) and in Twitter-only scores ($p<0.001$). There was not enough evidence to conclude a significant difference in lecture-only scores ($p=0.124$). Students who submitted answers to Twitter "all or most of the time" or "about half the time" had significantly higher overall scores and Twitter-only scores ($p<0.001$ and $p<0.001$, respectively) than those students who only submitted answers "little or none of the time."

Conclusion: While students retained less information from Twitter than from traditional classroom lecture, some retention was noted. Future research on social media in medical education would benefit from clear control and experimental groups in settings where quantitative use of social media could be measured. Ultimately, it is unlikely for social media to replace lecture in medical curriculum; however, there is a reasonable role for social media as an adjunct to traditional medical education. [West J Emerg Med. 2015;16(6):961–964.]

INTRODUCTION

The medical profession is one steeped in tradition. As many pre-medical hopefuls are informed – medicine is not only a career; it is a lifestyle dedicated to constant inquiry and the ascertainment of knowledge. Physicians are expected to grow and adapt to evolutions in their patients' environments. So, too, must the medical curriculum adjust to the millennial generation's affinity for technologically-driven mediums. In a 2011 study, Bosslet et al found that personal online social network (OSN) use among physicians and physicians-in-training was comparable to the general population; specifically, 90+% of medical students endorsed some form of OSN use,[1] with Facebook recognized as the most popular site on the Internet,[2] and Twitter reporting 302 million active users as of April 2015.[3] The advent of the social media age calls for a rapid adjustment in communication methods and curriculum goals. A 2013 systematic review by Cheston et al explored 14 studies' reports on the effects of social media in medical education. In their discussion, they stated that the use of social media in medical education merited further exploration, and suggested the benefit of more studies with "clear definitions of social media technologies… to allow appropriate comparisons and data synthesis.[4]

With these study deficits in mind, we sought to explore the use of a single social media focus (Twitter) and its effect on information exposure and retention within a second-year medical school class.

METHODS

During the 2013-2014 academic year, 116 second-year medical students were mandatorily enrolled in Introduction to Clinical Medicine II (ICM2) as part of their curriculum. All students enrolled in this course were considered responsible for its lecture content, per the syllabus distributed and reviewed on the first day of class. Within the course's weekly lecture series, images of clinically-relevant physical exam findings were presented to students. Outside the requirements of the class, students were invited to create a Twitter account and participate in an optional "Twitter Question of the Week." In this assessment, a short clinical scenario, accompanied by imaging of a relevant physical finding, was "tweeted out" weekly at pre-determined and publically-announced intervals. The twitter-associated clinical scenario and image were selected from content not covered in the ICM2 mandatory curriculum. Students were challenged to name the diagnosis based off the presented data. Student answers were submitted by direct messaging to the course instructor's account. The time stamp from each student's message enabled the instructor to identify the first 20 correct answers each week. These first 20 correct respondents were awarded "bonus points" (worth a fraction of a percent) to be added to their next ICM2 exam grade. Over the course of the year, over 80 students attempted an answer in at least one "Twitter Question of the Week."

After the completion of the course, students were asked to participate in a slideshow "quiz," with no effect on their grade. One hundred sixteen students (100% participation) were tested over their recognition of 24 clinical images—12 images from the lecture hours, and 12 images from the weekly twitter questions. These 24 images were assessed by a physician and third-year clinical medical student and selected for their straightforward, commonplace physical exam findings. Each image was projected for 30 seconds while students used pencil and paper to name the associated diagnosis or relevant exam finding. Course-based and Twitter-based photos were alternated in a 1:1 fashion.

On the same paper as their "quiz" answers, students were asked to complete a seven-point survey regarding each student's use of Twitter prior to the ICM course, his/her level of participation in the "Twitter Question of the Week" (did the student follow the ICM2 account? How many times did a student see/submit an answer to the twitter account? How many times did the student earn a "top 20" response?), and his/her reflections on educational Twitter usage (did the student feel like s/he retained more information regarding physical exam findings? Did this account enhance the student's overall education in ICM2?). Quiz/survey answers were then collected and analyzed by a statistician within the university.

RESULTS

A one-way analysis of variance (ANOVA) was conducted to determine the effect Twitter participation had on total scores, Twitter-only scores, and lecture-only scores. Twitter participation data was collected from the end-of-course survey and was defined as submitting answers to the Twitter-only questions "all or most of the time", "about half of the time", and "little or none of the time." The one-way ANOVA (Figure 1) determined that there was a significant difference in overall scores ($p<0.001$) and in Twitter-only scores ($p<0.001$), but there was not enough evidence to conclude a significant difference in lecture-only scores ($p=0.124$). A post hoc analysis using Tukey's range test (Figure 2) determined that those students who submitted answers to Twitter "all or most of the time" or "about half the time" had significantly higher overall scores ($p<0.001$ and $p<0.001$, respectively) and significantly higher Twitter-only scores ($p<0.001$ and $p<0.001$, respectively) than those students who only submitted answers "little or none of the time."

DISCUSSION

As the integration of social media into undergraduate and continuing medical education is still in its infancy, much of the literature has focused on descriptive terminology and anecdotal case studies. Generally, social medial integration into the curriculum has been well received by students and increases engagement, collaboration, and feedback, which may explain current trends to incorporate social media into medical education.[5-9] However, no substantial evidence exists

How often did you submit an answer on the twitter account?	Number of Students	Least Squares Mean	95% Confidence Interval		Tukey's HSD Test p-values		
			Lower	Upper	All of most of the time	About half the time	Little or none of the time
All or most of the time	23	11.696	10.469	12.922	-	.9212	<0.0001
About half the time	27	11.370	10.239	12.502	.9212	-	0.0001
Little or none of the time	66	8.333	7.609	9.057	<0.0001	0.0001	-

Figure 1. Least-squares means for all questions by answering questions on Twitter.
HSD, honest significant difference

How often did you submit an answer on the twitter account?	Number of Students	Least Squares Mean	95% Confidence Interval		Tukey's HSD Test p-values		
			Lower	Upper	All of most of the time	About half the time	Little or none of the time
All or most of the time	23	5.043	4.265	5.821	-	0.3453	<0.0001
About half the time	27	4.296	3.578	5.014	0.3453	-	<0.0001
Little or none of the time	66	2.106	1.647	2.565	<0.0001	<0.0001	-

Figure 2. Least-squares means for Twitter exclusive questions by answering questions on Twitter.
HSD, honest significant difference

to suggest that social media platforms are a comparable or superior form of medical education compared to current curriculums.[10-12] In light of these studies, as well as the results of our own research, we submit that while Twitter is not an appropriate replacement for aspects of medical education, it may have a role as an adjunct to traditional curriculums, as well as serve as another outlet to engage medical students on a more individual level. Within our research, students scored, on average, 54% of the course content questions correctly, and 27% of the Twitter questions correctly. Although their level of retention was not as high when compared to traditional classroom lectures, students who participated in the "Twitter Question of the Week" did retain some information from that medium, and a majority of the students in the class felt that Twitter added to their education.

This research study has several limitations. By declaring involvement in the Twitter project optional, the study risks self-selecting students who are already more academically-engaged into the Twitter-participant group. Alternatively, because the Twitter quizzes did not count towards ICM grades, students may have been less likely to return to and review the material outside of that week's presentation. Additionally, the slideshow quiz at the end of the course used the exact photos from Twitter questions, allowing for the possibility that students simply recognized the photo at hand, rather than interpreted the findings and demonstrated understanding of the medical condition.

In light of these limitations, future research on social media's role in medical education would benefit from clear control and experimental groups in settings where quantitative use of social media could be measured. If this study were to be replicated at other sites, it would be beneficial to use other photos/presentations of the selected exam findings to more accurately measure students' comprehension of the diseases. Ultimately, it is unlikely for social media to replace lecture within the medical curriculum; however, there is a reasonable role for social media as an adjunct to traditional medical education.

Address for Correspondence: Amanda L. Webb, MS, University of Kentucky School of Medicine, 220 Cedar Street, Apartment 109, Lexington, KY 40508. Email: alwe236@uky.edu.

Conflicts of Interest: By the *West*JEM article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. The authors disclosed none.

REFERENCES

1. Bosslet GT, Torke AM, Hickman SE, et al. The patient-doctor relationship and online social networks: results of a national survey. *J Gen Intern Med*. 2011;26(10):1168-74.
2. Hitwise (2010) Top 20 sites and engines. Available at: http://www.hitwise.com/us/datacenter/main/dashboard-10133.html. Accessed May 3, 2011.
3. Twitter About. 'Twitter Reports First Quarter 2015 Results; Lowers Full-Year 2015 Expectations'. N.p., 2015. Web. Accessed Sept 25, 2015.
4. Cheston CC, et al. 'Social Media Use In Medical Education: A Systematic Review. - Pubmed - NCBI'. Available at: Ncbi.nlm.nih.gov. N.p., 2015. Web. Accessed Sept 25, 2015.

5. Mckay M, Sanko JS, Shekhter I, et al. Twitter as a tool to enhance student engagement during an interprofessional patient safety course. *J Interprof Care*. 2014;28(6):565-7.
6. Ravindran R and Vivekananthgam S. Harnessing social media for medical education. *Clin Teach*. 2014;11:237-9.
7. Bahner D, Adkins E, Patel N, et al. How we use social media to supplement a novel curriculum in medical education. *Med Teach*. 2012;34:439-44.
8. Jalalat S and Wagner R. Utility of a dermatology interest group blog: the impact of medical student interest groups and web 2.0 tools as educational resources. *Adv Med Educ Pract*. 2014;5:331-7.
9. George D and Dellasega C. Social media in medical education: two innovative pilot studies. *Med Educ*. 2011;45(11):1158-9.
10. Cartledge P, Miller M, Phillips B. The use of social-networking sites in medical education. *Med Teach*. 2013;35:847-57.
11. Pander T, Pinilla S, Dimitriadis K, et al. The use of Facebook in medical education–a literature review. *Z Med Ausbild*. 2014;31(3):1-19.
12. Cheston CC, Flickinger TE, Chisolm MS. Social media use in medical education: a systematic review. *Acad Med*. 2013;88:893-901.

SAVE THE DATE

CDEM CONSENSUS CONFERENCE ON CLINICAL ASSESSMENT OF MEDICAL STUDENTS

March 6-7, 2016

CORD Academic Assembly, Nashville, TN

Program Directors, Clerkship Directors, Deans, Students, and Residents:

We want your input on how students are assessed during their EM clerkship. Join us for the consensus conference at the CORD Academic Assembly in Nashville, TN to develop a national assessment tool for students. Attendees will participate in discussions happening now around major themes including:

- Criterion- vs. norm-referenced assessment
- Learners at different levels
- Translation of clinical assessment data into other products
- Faculty/resident development and implementation
- Ensuring post-implementation validation of a national tool

Learn more at www.cordem.org

Not a member of CDEM? Consider joining & contributing to medical student education. Join at www.saem.org.

Emergency Department Psychiatrist

In partnership with Regions Hospital Emergency Department and Behavioral Health nursing staff, crisis social workers and Emergency Medicine providers, this position will lead the coordination, quality and efficiency of acute care for mental health patients at their point of presentation to Regions Hospital's Emergency Department (ED) in St. Paul, Minnesota. In addition to being a top Upper Midwest Level 1 trauma and burn center, Regions Hospital is a national award-winning, innovative provider of Behavioral Health services to the Minneapolis/St. Paul metropolitan area and western Wisconsin. The department sees 80,000 visits per year with 10% mental health emergencies.

We seek a BC/BE Psychiatrist who is flexible, creative and engaging; has the ability to triage multiple clinical possibilities with limited datasets under competing demands on skills and time; possesses excellent interpersonal communication skills; and is a strong clinician with a mature sense of priorities/practical experience who can formulate and implement treatment pathways within the framework of hospital, medical group and community resources. Experience in psychosomatics and addictions with an aptitude for liaison work would be valuable assets, and the ability to guide multidisciplinary trainees with teaching and supervision in the emergency environment is a must. This full-time position involves direct patient care in a consultative role within Regions Hospital's ED and includes participation in our related network referral service and coordination with community-based providers to ensure the best continuity of care for emergency mental health patients.

An outstanding benefits and compensation package, talented and energetic staff/colleagues, a rewarding practice and an exciting metropolitan lifestyle await! Submit your CV to **lori.m.fake@healthpartners.com** or complete our online application at **healthpartners.com/careers**. For more details, please contact Lori Fake at 952-883-5337 or 800-472-4695 x1. EOE

HealthPartners®

2016 Emergency Medicine in Yosemite

January 13 – 16, 2016

Sponsored by Yosemite Medical Education Foundation (YMEF)

Co-Sponsored by California ACEP

(12.0 *AMA PRA Category 1 Credits*™)

"This activity has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education through the joint provider ship of the Center for Emergency Medical Education and Yosemite Medical Education Foundation."

"The Center for Emergency Medical Education is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians."

The Center for Emergency Medical Education designates this live activity for a maximum of 12.00 *AMA PRA Category 1 Credits*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Approved by the American College of Emergency Physicians for a maximum of 12.00 hour(s) of ACEP Category I credit.
Approved for 12.0 credits of AOA Category 2-A credits

Physicians should claim only the credit commensurate with the extent of their participation in the activity.

CORD EM

AND

EMF Emergency Medicine Foundation

research • education • patient care

Announce two new grants for the 2016 Grant Cycle!

Deadline to Apply: February 12, 2016

- **CORD-EMF Starter Grant**
- **CORD-EMF Education Research Grant**

For more information or to apply for a grant please visit:

www.cordem.org/grants

Ohio ACEP
American College of Emergency Physicians
Advocacy | Education | Leadership

OHIO ACEP IS EXCITED TO ANNOUNCE A DATE & LOCATION IN PARTNERSHIP WITH CALIFORNIA ACEP!

CALIFORNIA ACEP
AMERICAN COLLEGE OF EMERGENCY PHYSICIANS

February 8 - 12, 2016
Newport Beach, CA

Ohio ACEP Emergency Medicine Board Review Course

The Ohio ACEP Emergency Medicine Board Review

February 8 - 12, 2016

Course Location
Newport Beach Marriott Hotel & Spa
Newport Beach, California

Approved for *AMA PRA Category 1 Credits ™*.

www.ohacep.org **(614) 792-6506**

COMPREHENSIVE.
RELEVANT.
ESSENTIAL.

CORD EM HEADS TO Nashville MUSIC CITY

Academic Assembly
March 6-9, 2016

Omni Nashville

Nashville, TN

New Expanded Tracks, CPC Semi-Final Competition, New Electronic Poster Session and Many more "Don't Miss Events"

visit www.cordem.org for details