AEM Education and Training
A GLOBAL JOURNAL OF EMERGENCY CARE

Elevating the human condition during times of emergency

ORIGINAL CONTRIBUTIONS

Applicant Attitudes Toward the Association of American Medical Colleges’ Standardized Video Interview
Ashlea D. Winfield, Neeraj Chhabra, Michael A. Schindlbeck et al. 5

Dissecting the Contemporary Clerkship: Theory-based Educational Trial of Videos Versus Lectures in Medical Student Education
Stella H. M. Yiu, Alena M. Spacek, Paul G. Pageau et al. 10

Does Applicant Gender Have an Effect on Standardized Letters of Evaluation Obtained During Medical Student Emergency Medicine Rotations?
Jessica Andrusaitis, Clelia Clark, Soheil Saadat et al. 18

The Correlation Between Emergency Medicine Residents’ Grit and Achievement
Adriana Segura Olson, Kelly Williamson, Nicholas Hartman et al. 24

Salivary Cortisol Concentrations, Grit, and the Effect of Time
Matthew L. Wong, Gregory Peters, Joshua W. Joseph et al. 30
The Equivalence of Video Self-review Versus Debriefing After Simulation: Can Faculty Resources Be Reallocated?
Gregory J. Tudor, Gregory S. Podolej, Ann Willemsen-Dunlap et al. 36

Defining “Swarming” as a New Model to Optimize Efficiency and Education in an Academic Emergency Department
Jessica L. Perniciaro, Anita R. Schmidt, Phung K. Pham et al. 43

Showing Your Thinking: Using Mind Maps to Understand the Gaps Between Experienced Emergency Physicians and Their Students
Kira Gossack-Keenan, Kerstin De Wit, Emily Gardiner et al. 54

BRIEF CONTRIBUTION

It’s Just Math—Unless It’s Toxic!
Kayla Myers, Elisabeth Giblin, Michele Zell-Kanter 64

INNOVATIONS REPORT

Successful Implementation of a Resident Liaison to Medical Students in Emergency Medicine Rotations
Jessica Bod, Alina Tsyrulnik, Ryan Coughlin et al. 68

EDUCATIONAL DOWNLOAD

Five Tips for Building a Successful Didactic Talk
Eric Steinberg, Doug Franzen 72

COMMENTARY–INVITED

In Reply to: Applicant Attitudes Toward the Association of American Medical Colleges’ Standardized Video Interview
Nicole M. Deiorio, Dana Dunleavy, Christopher M. Woleben 75

COMMENTARY AND PERSPECTIVE

Gender-based Harassment of Emergency Medicine Trainees: What Faculty Educators Need to Know
Dawn Jung 77
Copyright and Copying (in any format)
Copyright © 2020 Society for Academic Emergency Medicine. All rights reserved. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA, www.copyright.com, provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kinds of copying such as copying for general distribution, for advertising or promotional purposes, for republication, for creating new collective works or for resale. Permissions for such reuse can be obtained using the RightsLink “Request Permissions” link on Wiley Online Library. Special requests should be addressed to permissions@wiley.com

Disclaimer
The Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors cannot be held responsible for errors or any consequences arising from the use of information contained in this journal; the views and opinions expressed do not necessarily reflect those of the Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors, neither does the publication of advertisements constitute any endorsement by the Publisher, the Society for Academic Emergency Medicine (SAEM) and Editors of the products advertised.

For submission instructions, subscription and all other information visit http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)2472-5390
Applicant Attitudes Toward the Association of American Medical Colleges’ Standardized Video Interview

Ashlea D. Winfield, MD, Neeraj Chhabra, MD, Michael A. Schindbeck, MD, and Steven H Bowman, MD

ABSTRACT

Background: The Standardized Video Interview (SVI) was developed by the American Association of Medical Colleges to allow applicants to include objective data about professional behaviors and interpersonal and communication skills. Although the SVI pilot was administered to individuals applying to emergency medicine (EM) residency programs during the 2018 Electronic Residency Application Service (ERAS) cycle, little data have been published evaluating the applicant’s perceptions. This survey aims to assess EM residency applicant attitudes toward the SVI.

Methods: During the 2018 ERAS application season an anonymous survey was administered to interviewees at one urban Accreditation Council for Graduate Medical Education-approved EM residency. Respondents were asked questions regarding the production of their video interviews, thoughts regarding the additive value of the SVI, and individual demographic data such as ethnicity and sex. Participation was optional.

Results: A total of 219 of 238 candidates completed the survey representing a 92% response rate. While the majority of applicants did not feel that their ethnicity impacted their application, 58.1% of those who did self-identified as African American or Asian. A total of 8.7% of respondents felt the SVI added information about their professional behaviors and 11% felt that it added information about interpersonal and communication skills. Only 2.8% of survey respondents felt the SVI should remain a portion of the ERAS application.

Conclusions: Most respondents felt that the SVI was not an accurate representation of their interpersonal and communication skills or their professionalism and that it did not add value to their applications. While most cohorts were not concerned about bias regarding sex, ethnicity, sex, or age, a small subset felt that there was a potential for the SVI to bias the party reviewing their applications. Very few applicants felt the SVI should remain a part of the ERAS application. Applicant attitudes toward the SVI are largely negative and require further investigation prior to becoming a standard part of applicants’ ERAS files.
Graduate Medical Education (ACGME) competencies that may not be readily discernable in other areas of an applicant’s file, namely, their knowledge of professional behaviors along with their interpersonal and communication skills. The SVI and a composite score were viewable by EM residency programs with each candidate’s ERAS application.

The SVI consists of six questions presented online via text prompts. The questions are not related to clinical knowledge and focused on subcompetencies such as emotional intelligence, ethics, empathy, and cultural competence. For example, “describe a time you worked with a challenging patient. What was the situation? What actions did you take?” The applicants have no prior knowledge of the questions and are given 30 seconds to read and reflect followed by a 3-minute period to record a video response. The video responses are produced at the applicant’s discretion at no increased cost to them. Modalities used included personal computers, electronic tablets, or smartphones.

Applicants have one opportunity to answer each question. Their responses are subsequently rated on a standardized scale developed specifically for the SVI and assigned a score between 1 and 5. With a total of six questions, cumulative scores range from a low of 6 to a high of 30, with a mean (±SD) of 18.7 (±2.8) per the 2017 AAMC pilot data. Six raters are assigned per interview and a different rater is assigned to each question with the goal of reducing any one rater influence on the overall score. All participating raters were selected from a pool of individuals “experienced in human resources and assessment.” Selected individuals then completed the AAMC’s unconscious bias training, are educated on EM resident job duties, and utilize a standardized method to evaluate responses. Following 12 hours of in-person training, raters practice rating applicants and receive direct feedback on their performance. Raters complete a total of 20 hours of training before being allowed to formally score applicants.

It is likely the SVI will become a standard part of applicants’ ERAS files for all specialties. While the SVI was developed by the AAMC to provide information to residency programs regarding the aforementioned competencies, there is limited published literature regarding the viewpoint of those most affected by its implementation, i.e., the applicants themselves. Some have called for a slowdown of SVI implementation until applicant representation in the SVI process is established. Others have questioned whether the SVI can provide an adequate measure of professionalism. We sought to assess applicant attitudes toward the SVI pilot at one EM residency program as well as information regarding the applicants’ production of their individual SVIs.

**METHODS**

This was a survey study of applicants at one large urban public hospital’s ACGME-approved EM residency program. The survey was administered to all applicants who underwent in-person interviews for the residency program’s 2018 application cycle. A survey instrument was used to ensure anonymity and minimize response modification by the applicants in the desire to garner a more favorable outcome during their interview day.

The survey was developed by the authors after a literature review determined that a relevant external survey had yet to be published. The survey was edited by faculty for clarity and relevancy. The survey was subsequently piloted to current EM resident physicians to determine if the survey was clear, minimized bias, and void of leading questions to increase construct validity. It was then presented to current medical students who were not participating in the upcoming ERAS application season to determine relevancy and ease of interpretation. The survey was separated into the following sections: SVI production; preparation; content; scoring; and demographic information such as race, sex, and ethnicity. The above domains were evaluated to gather general data on preparation and production modalities, to assess applicants’ perceived utility of the SVI, applicants’ knowledge of the scoring process, and perceived bias (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10355/full).

The survey was administered during the 11 interview days of the 2018 application cycle spanning from October 17, 2017, to January 30, 2018. It was completed at the end of the interview day to minimize the potential confounder of an applicant attempting to answer survey items in a way that could be perceived as beneficial to his or her interview process. Subjects were allowed to answer however many of the questions as they wished and were allotted as much time as needed to respond, and all answered questions were included in the data. Survey data were subsequently abstracted into a computerized spreadsheet by trained research assistants. The study was approved by the local institutional review board. Additionally, the
AAMC was contacted and permission was given to complete the study.

RESULTS

A total of 219 of 238 applicants completed some portion of the survey representing a 92% response rate. Demographic details of respondents are presented in Table 1. Self-reported SVI scores ranged from 10 to 29 with a median score of 20. SVI production modalities were fairly consistent among responders with most using a laptop computer to produce their SVI (193/219, 88.1%) which was filmed in their personal or private residence (151/219, 68.9%). A minority of respondents produced their SVI in a medical school facility (56/219, 25.6%). Technical difficulties were noted by 16 of 219 respondents (7.3%), and only three of 219 respondents reported that their medical school used a professional video production company to complete their SVI (1.4%).

Table 2 represents applicant attitudes toward the SVI. Overall 32% (70/219) of applicants were satisfied with preparation for the SVI by the AAMC, 42% (96/219) felt that their medical schools offered adequate preparation. (96/219, 42%). A total of 22.8% (50/219) applicants felt they were understood the scoring process. While some respondents reported concerns that characteristics such as age, sex, sex, ethnicity, and appearance had an effect on their SVI scores, 75.5% to 92% felt that there was no effect for the listed categories. Although AAMC data from the 2017 to 2018 SVI administration showed that there were no scoring differences along race or ethnicity, 58.1% of applicants expressing concern over a negative effect of their ethnicity on their SVI score self-identified as either African American or Asian (Table 3).

DISCUSSION

The SVI was developed to provide objective data in respect to applicants’ professional behaviors along with their interpersonal and communication skills with the overall goal to add depth to their applications and

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Respondents (n = 219)</th>
<th>AAMC 2017 Pilot Data (n = 1,760)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (n = 205)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>86 (42.0)</td>
<td>734 (39.5)</td>
</tr>
<tr>
<td>Male</td>
<td>119 (58.0)</td>
<td>516 (60.5)</td>
</tr>
<tr>
<td>Age, years (n = 202)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>27 (26–28)</td>
<td>NR</td>
</tr>
<tr>
<td>Range</td>
<td>24–42</td>
<td>NR</td>
</tr>
<tr>
<td>Race/ethnicity (n = 205)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>2 (1.0)</td>
<td>NR</td>
</tr>
<tr>
<td>Asian</td>
<td>31 (15.1)</td>
<td>455 (25.6)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>30 (14.6)</td>
<td>116 (10.6)</td>
</tr>
<tr>
<td>Hispanic, Latino, of Spanish origin</td>
<td>23 (11.2)</td>
<td>126 (7.4)</td>
</tr>
<tr>
<td>White</td>
<td>109 (53.2)</td>
<td>921 (52.6)</td>
</tr>
<tr>
<td>Multiple</td>
<td>4 (2.0)</td>
<td>NR</td>
</tr>
<tr>
<td>Other</td>
<td>6 (2.9)</td>
<td>NR</td>
</tr>
</tbody>
</table>

Data are reported as n (%) unless otherwise reported. AAMC = American Association of Medical Colleges; IQR = interquartile range; NR = not reported.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Yes</th>
<th>No</th>
<th>Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production and preparation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Were you satisfied with the technical quality of your SVI?</td>
<td>149 (68.0)</td>
<td>25 (11.4)</td>
<td>45 (20.5)</td>
</tr>
<tr>
<td>Were you provided with adequate preparation by the AAMC to answer the questions of the SVI?</td>
<td>149 (68.0)</td>
<td>25 (11.4)</td>
<td>45 (20.5)</td>
</tr>
<tr>
<td>Did your medical school provide you with any preparation and/or coaching about the SVI?</td>
<td>149 (68.0)</td>
<td>25 (11.4)</td>
<td>45 (20.5)</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did your SVI reflect additional information on your knowledge of professional behaviors not available elsewhere in your ERAS application?</td>
<td>19 (8.7)</td>
<td>174 (79.5)</td>
<td>26 (11.9)</td>
</tr>
<tr>
<td>Did your SVI reflect additional information on your interpersonal and communication skills not available elsewhere in your ERAS application?</td>
<td>24 (11.0)</td>
<td>171 (78.4)</td>
<td>23 (10.6)</td>
</tr>
<tr>
<td>If you had the ability to redo any of your answers, would you change any?</td>
<td>107 (48.9)</td>
<td>46 (21.0)</td>
<td>66 (30.1)</td>
</tr>
<tr>
<td>Scoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you feel adequately informed of the scoring process prior to filming your SVI?</td>
<td>50 (22.8)</td>
<td>156 (71.2)</td>
<td>13 (5.9)</td>
</tr>
<tr>
<td>Should the SVI remain a part of the ERAS application?</td>
<td>6 (2.8)</td>
<td>187 (86.2)</td>
<td>24 (11.1)</td>
</tr>
</tbody>
</table>

Data are reported as n (%). AAMC = American Association of Medical Colleges; ERAS = Electronic Residency Application Service; SVI = Standardized Video Interview.
potentially increase the number of interviews granted to applicants that may not have otherwise been consid-
ered. If the SVI can serve to identify additional profi-
cencies in these competencies, it is likely that its use will expand to residency applications across all medical specialties. Unfortunately, there have not been many studies evaluating the perspective of those most affected by the implementation of the Standardized Video Interview, the residency applicants. This study is one of the first to query applicants who have completed the SVI on their production modalities and their attitude toward its use in ERAS.

The overarching goal of the SVI was to provide objective data not available elsewhere in an ERAS application. To this end, only 11% of applicants felt that the SVI accomplished this goal.

The majority of respondents felt that they were insufficiently prepared to answer the SVI questions by either the AAMC or their medical schools and that they lacked an adequate knowledge of the scoring process. It is not clear what preparatory materials were given to applicants prior to implementation of the SVI. Furthermore, most of them would have revised their answers had they been given the opportunity to do so. These concerns should be addressed by both the AAMC and the applicants’ medical schools. Additionally, while the majority of respondents were unconcerned about scoring biases related to their age, sex, and gender, a minority expressed concern regarding the impact of video production value, ethnicity, and personal physical appearance on their SVI score. Only 2.8% (6/219) of respondents reported that the SVI should remain a part of the ERAS application. This could be an understandable response to any and all of the following factors: yet another requirement being placed on the shoulder of applicants, frustration with the perceived lack of transparency in scoring, the underlying potential for bias, and a concern that the Standardized Video Interview does not add value to their application.

While the data presented in this study do provide a snapshot of applicant attitudes toward the SVI, they are inherently limited due to selection bias. This study was conducted at only one ACGME-approved EM residency program. Additionally, only applicants that were selected to interview were included in the study that potentially excludes applicants with lower SVI scores. Furthermore, although steps were taken to minimize potential confounders, including making the survey anonymous and administering it after the interview day was complete, the potential for self-reporting bias still exists. Additionally, the authors were unable to follow-up with respondents due to the anonymity of the process, but this was felt best as it allowed more honest reporting by applicants.

As residency programs prepare their rank lists, there is always a leap of faith in interpreting how an applicant’s past performance and in-person interview will translate to residency performance. Utilizing additional information to bridge that gap is a worthy endeavor; however, adding additional work on the applicants’ end, when they are already performing clinical rotations, studying clinical content, paying for medical school, and paying for their interview process, may be intrinsically unfair to them if they do not identify a clear benefit or, even worse, perceive the potential introduction of bias.

In March 2019 Academic Medicine published an article evaluating the validity of the Standardized Video Interview scores. There were two studies. Study 1 (2016 cohort) included 855 applicants applying to EM, pediatrics, and internal medicine. Study 2 (2017 cohort) included 3,532 applicants applying to EM during the 2018 ERAS cycle. It was noted in their article that there were none-to-small correlations between SVI scores and unrelated academic variables such as Step scores, implying that the SVI is “measuring something different than academic performance.” While this may be true, further research is needed to ensure that the SVI score is an accurate representation of professional behaviors and interpersonal and communication skills. This study does not expel applicant concern that the Standardized Video Interview may not accurately reflect these domains.

Table 3
Applicant Concerns Regarding Bias on the SVI

<table>
<thead>
<tr>
<th>What Effect, If Any, Do You Think the Following Personal or SVI Characteristics Had on the Score of Your SVI?</th>
<th>Negative</th>
<th>No Effect</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (n = 196)</td>
<td>9 (4.6)</td>
<td>180 (91.9)</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>Sex (n = 193)</td>
<td>15 (7.8)</td>
<td>167 (86.5)</td>
<td>11 (5.7)</td>
</tr>
<tr>
<td>Gender (n = 194)</td>
<td>18 (9.3)</td>
<td>164 (84.5)</td>
<td>10 (5.2)</td>
</tr>
<tr>
<td>Ethnicity (n = 194)</td>
<td>31 (16.0)</td>
<td>147 (75.7)</td>
<td>16 (8.2)</td>
</tr>
<tr>
<td>Physical appearance (n = 194)</td>
<td>28 (14.4)</td>
<td>146 (75.3)</td>
<td>20 (10.3)</td>
</tr>
<tr>
<td>Video production value (n = 196)</td>
<td>50 (25.5)</td>
<td>133 (67.9)</td>
<td>13 (6.6)</td>
</tr>
</tbody>
</table>

Data are reported as n (%).
SVI = Standardized Video Interview.
CONCLUSION

Despite the stated goal of providing an additional measurement of competencies for residency applicants that may be lacking in the current ERAS format, applicant attitudes toward the SVI are largely negative. These attitudes need to be respected and criticisms fully explored before the expansion of the pilot to a larger scale. Further work is needed to determine whether the SVI helps programs via allowing objective measures of constructs not typically available, adds an unnecessary burden to the application process, or potentially introduces biases that may hurt a select subset of applicants.

The authors acknowledge the Academic Associates Program of the Department of Emergency Medicine of Cook County Hospital for their assistance in enrolling study subjects and abstracting data. Specifically, we acknowledge the roles of the Research Associates Program Coordinator, Lum Rizvanolli, and Clinical Research Coordinator, Errick Christian.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10355/full

Data Supplement S1. Standardized Video Interview (SVI) Questionnaire.
Dissecting the Contemporary Clerkship: Theory-based Educational Trial of Videos Versus Lectures in Medical Student Education

Stella H. M. Yiu, MD, MEd1,2, Alena M. Spacek, MD1, Paul G. Pageau, MD1,2, Michael Y. C. Woo, MD1,2, A. Curtis Lee, PhD3,5, and Jason R. Frank, MD, MA(Ed)5

ABSTRACT

Background: Despite increasing use of the flipped classroom (FC) technique in undergraduate medical education, the benefit in learning outcomes over lectures is inconsistent. Best practices in preclass video design principles are rarely used, and it is unclear if videos can replace lectures in contemporary medical education.

Methods: We conducted a prospective quasi-experimental controlled educational study comparing theory-based videos to traditional lectures in a medical student curriculum. Medical students enrolled in an emergency medicine clerkship were randomly assigned to either a lecture group (LG) or a video group (VG). The slide content was identical, and the videos aligned with cognitive load theory-based multimedia design principles. Students underwent baseline (pretest), week 1 (posttest), and end-of-rotation (retention) written knowledge tests and an observed structured clinical examination (OSCE) assessment. We compared scores between both groups and surveyed student attitudes and satisfaction with respect to the two learning methods.

Results: There were 104 students who participated in OSCE assessments (49 LG, 55 VG) and 101 students who participated in knowledge tests (48 LG, 53 VG). The difference in OSCE scores was statistically significant 1.29 (95% confidence interval = 0.23 to 2.35, t(102) = 2.43, p = 0.017), but the actual score difference was small from an educational standpoint (12.61 for LG, 11.32 for VG). All three knowledge test scores for both groups were not significantly different.

Conclusions: Videos based on cognitive load theory produced similar results and could replace traditional lectures for medical students. Educators contemplating a FC approach should devote their valuable classroom time to active learning methods.

Medical educators have increasingly proposed using a flipped classroom (FC) for medical student education.1,2 In this model, the traditional lecture is a student’s preclass homework, while in-class time is spent on active, inquiry-based learning facilitated by an instructor.1,3–5 Most preclass lectures were recorded in the forms of podcasts, screencasts, and videos.6 Despite the promise of FC to promote active learning, outcomes are inconsistent.6–8 By dissecting how each element (preclass and in-class) contributes to learning,
perhaps we can design a better medical student curriculum. The FC classroom incorporates active learning, with teachers explaining in-depth knowledge instead of merely providing factual information, and students spend time discussing with peers.9 Apart from being interactive,8,10 there is much heterogeneity in how in-class activities are conducted, and little is known about how specific in-class activities contribute to learning.7

When we focus on preclass videos, literature comparing videos to lectures is unclear, with some studies favoring videos,1,2,11–14 and some finding no difference in learning outcomes.12,14–17 In other studies, video groups showed worse learning outcomes,3,12,18–21 with authors citing reasons including no peer interaction, low compliance with learning plans, and lack of accountability.1,3–5 Studies reported suboptimal use of videos,7 with view rate from around 60%22–24 to 83%.25 With no clear evidence to support videos, some authors have urged educators to stop replacing lectures with them.26–28

If we dissect these videos further, the variation in learning outcomes might be due to differences in design.29 Few curriculum developers subscribed to guidelines guided by pedagogical principles.29,30 As there is diverse configuration, methods, and presentation,31 Mayer’s multimedia design principles based on cognitive load theory have been espoused to guide future designs.33–35 In Mayer’s approach, multimedia instruction has the potential for cognitive overload, as the learner’s cognitive processing exceeds the available capacity.34,36 His principles reduce cognitive load to maximize learning.22,36–38 However, with few exceptions, most studies did not explicitly reference any design.36,39,40 Scant description of video production was typically limited to the software programs and hardware equipment used. Format, if mentioned, varied.27 Examples included interactive games,31 animations,33 or videotaped lectures.22,37 Given the lack of detail in these preclass materials, study replication is difficult.

Therefore, while the in-class components are important, it is possible that video design partly accounts for the variability in published FC outcomes. To compare and replicate studies, it would be useful for preclass videos to adhere to good practice for format and design. More specifically, we need to establish whether videos adhering to Mayer’s principles could replace lectures in the FC model.

We set out to determine if educational videos adhering to Mayer’s multimedia design principles resulted in any difference in learning outcomes for medical students when compared to traditional lectures in the setting of an emergency medicine undergraduate clerkship.

**METHODS**

We conducted a quasi-experimental controlled study to evaluate the effectiveness of videos compared with lecture for teaching medical students trauma assessment.

**Study Setting and Population**

All 113 third-year English stream medical students entering their mandatory emergency medicine rotation from the university were invited to participate and enrolled in a continuous fashion from March 2013 to March 2014. We received approval from the research ethics board from the Ottawa Research Health Institute and the University of Ottawa. As per institutional review board–approved protocol, students were informed that their participation and results would not affect their rotation assessment.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Examples in Slides Using Mayer’s principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principle</td>
<td>Coherence principle: eliminate extraneous material</td>
</tr>
<tr>
<td>Examples</td>
<td>One to two keywords only: “airway assessment,” “pneumothorax.”</td>
</tr>
<tr>
<td>Principle</td>
<td>Multimedia principle: present words and pictures rather than words alone</td>
</tr>
<tr>
<td>Examples</td>
<td>On “mechanism” of trauma, a single photo of a wrecked car is shown.</td>
</tr>
<tr>
<td>Principle</td>
<td>Signaling principle: highlight essential material</td>
</tr>
<tr>
<td>Examples</td>
<td>Included intro and summary. Slide heading emphasized key points.</td>
</tr>
<tr>
<td>Principle</td>
<td>Contiguity principle: place printed words near corresponding graphics</td>
</tr>
<tr>
<td>Examples</td>
<td>On slide with tension pneumothorax keywords “tension pneumothorax.”</td>
</tr>
<tr>
<td>Principle</td>
<td>Pretraining principle: provide pretraining in names and characteristics of key concepts</td>
</tr>
<tr>
<td>Examples</td>
<td>On primary survey slide included the list of “airway, breathing, circulation, disability and exposure.”</td>
</tr>
</tbody>
</table>
Lecture and Video Development
An emergency medicine clerkship director created the content for an “approach to trauma” presentation with university curriculum objectives. The PowerPoint (Microsoft Corp.) slide design adhered closely to Mayer’s principles \(^{38}\) (see Table 1). There was minimal text (coherence principle). Instead, there was relevant nonanimated graphics with a few words next to them (multimedia principle). Each section started with a highlighted introduction and summary (signaling principle). These slides formed the basis of the lecture and the videos. The planned content was written as a script.

The same instructor used the slides and script to record five voice-over videos (spanning 6–12 min). The videos further engaged Mayer’s personalization and voice principles. Specifically, content was segmented into five videos (segmenting principle). Video contained narrated script (modality principle) in a conversational style (personalization principle) by a person (voice principle). The videos also use contiguity principle by showing words at the exact time of narration placed near corresponding graphics (see Figure 1).

The slides and videos were pilot-tested by another clerkship director and students. After suggested modifications, the videos were uploaded to a password-protected Web page. An embedded Google Analytics tracker logged visits anonymously.

In the lecture, the instructor presented the slides over a 2-hour session with the planned script. The first lecture was recorded and audited by another instructor to ensure consistency with the planned script.

Study Protocol
Once consent was obtained, participants were enrolled in the study. The university had already randomly assigned students into groups (of either 15 or 16) prior to start of clerkship. These groups were created to rotate student through the various disciplines during clerkship training. These groups were then assigned to either lecture group (LG) or video group (VG) in an alternating fashion.

On day 3, LG students attended the lecture and VG students were given access to the videos. On day 8, each participant performed alone in an observed structured clinical examination (OSCE) in a high-fidelity simulation scenario in trauma assessment and management. Each performance was recorded. Participants also completed three written knowledge tests on day 1 (pretest), day 8 (posttest), and week 5/6 (retention test). After data collection, LG students were also given access to the videos.

Measurements. Our primary outcome measure was application of trauma management skills. Secondary measures included acquired knowledge and attitudes to the learning methods used.

OSCE. The OSCE assessment consisted of a checklist (total 20) and an anchored global rating scale.\(^{31}\) We delineated three domains of trauma management (airway, breathing, and circulation) and identified specific objective elements for each domain. Each item on the checklist requires an application of the cognitive domain of the objective. Each performance was recorded. Three raters (staff emergency physicians) independently reviewed two pilot videos and were trained on the checklist. After that, they discussed and revised unclear items on the checklist through consensus. These raters all had 5 or more years of experience rating OSCE performances. Two independent raters scored each performance on a later date. These raters were blinded to the group allocation and each other’s score. There were three raters—rater 1 scored all recordings, with rater 2 and 3 scoring 50% of recordings, respectively.

Knowledge Assessments. We created 60 multiple-choice questions based on representative sample of items in varying difficulty linked to the content domain. We followed previous guidelines for question construction.\(^{42,43}\) The questions matched the learning objectives of the videos and the lecture. They were
reviewed by three expert content reviewers (staff emergency physicians) and piloted on a separate group of students before the study. We refined ambiguous wordings. The questions were randomly divided based on the main content domains tested into three 20-item tests. No interitem correlation for the content domains was analyzed. The participants answered these three knowledge tests on day 0, day 8, and week 5 or 6 depending on the rotation schedule.

Survey. The survey assessed the students’ experience and attitudes toward their assigned modality and learning style prior to the simulation. The question “Would you have preferred what you were assigned to?” was repeated after the OSCE simulation assessment. A research assistant extracted and collected the data. A random 10% of sample data was reviewed for accuracy.

Data Analysis
Study data were analyzed using SAS software. The experts agreed by consensus that a 10% difference in the OSCE score (an absolute value of 2) was educationally significant, in keeping with prior studies. Assuming a normal distribution, 100 participants (50 in each group) would provide a power of 0.80 to detect a 10% difference between the groups (two-sided test at \( \alpha = 0.050 \)). Differences in examination scores between groups and their two-sided 95% confidence intervals (CIs) were calculated using Student’s t-tests. Cohen’s \( \delta \) was calculated for practical significance. Inter-rater reliability was calculated between the independent raters. Categorical qualitative data were analyzed using Fisher’s exact test. For ordinal data in the surveys (Likert scales), nonparametric statistics were used. Participants were analyzed in the groups they were assigned to.

RESULTS
All 113 students entering the emergency medicine rotation provided consent. Half of these groups (two VG, two LG) have had prior surgery rotations while the rest (two VG, two LG) have not. They all had prior OSCE experience. There were 104 recordings (49 LG, 55 VG) available for analysis (see Figure 2).

Primary Outcome
The means of the OSCE scores were 12.61 and 11.32 for the LG and VG, respectively. The mean difference
was 1.29 (95% CI = 0.23 to 2.35, t(102) = 2.43, p = 0.017), indicating no educationally significant difference (as defined a priori as > 2). Using an independent t-test (t(102) = 2.43), the difference in the OSCE scores was statistically significant (p = 0.017). Cohen’s d was 0.481, suggesting a moderate practical significance (see Figure 3). Inter-rater reliability was good between rater 1 and 2 (intraclass correlation [ICC] = 0.834, 95% CI = 0.685 to 0.909) and excellent between rater 1 and 3 (ICC = 0.953, 95% CI = 0.920–0.973).

Secondary Outcomes

Knowledge Tests. Test scores from 101 participants were available for analysis (48 LG, 53 VG). There were no statistically significant differences between the pretest, posttest, and retention scores for the two groups (see Figure 4). Cohen’s d was 0.105 and 0.034, respectively, indicating no practical difference.

Video Use. The highest numbers of visits occurred on the day prior to simulation (mean = 11.75, 10–16). Of the VG group, 90% (45 of 50) self-reported that they had viewed all the videos. There were 25% (13 of 52) students who reported repeated viewing. The number of page visits over the rotation totaled more than the number of students, also suggesting repeated visits. These visits also captured potential LG students who were given access to the videos after data collection.

Survey. There were 103 students who participated in the survey (51 LG, 52 VG). Most students agreed (somewhat or strongly) that there were clear learning goals (88% LG vs. 94% VG, p = 0.1562) and the correct level of complexity (90% LG vs. 94% VG, p = 0.4878) and organization (85% LG vs. 96% VG, p = 0.0923). The majority of students (73% LG vs. 65% VG) agreed (somewhat or strongly) that it was better than the average teaching session (p = 0.5754). Significantly more LG (80%) than VG (33%) agreed (somewhat or strongly) that the session encouraged collaboration (p < 0.0001) and was interactive (88% LG vs. 34% VG, p < 0.0001) but was more demanding (28% LG vs. 13% VG, p = 0.0109). Most students agreed (somewhat or strongly) that the session was flexible to meet their needs (82% LG vs. 85% VG, p = 0.5195) and answered their own questions (85% LG vs. 73% VG, p = 0.3348).

Learning Preferences. Significantly more LG (71%) than VG (37%) preferred learning with a lecturer (p = 0.002), while more VG (73% VG vs. 55% LG, p = 0.1216) stated that they preferred learning on their own time. A third of both groups (33% LG vs. 34% VG, p = 0.9659) stated that they were prompted to look up their own resources.

Application. Most students (75% LG vs. 81% VG, p = 0.4248) felt that the session would help them apply their knowledge in a case scenario. Prior to simulation, 81% of the LG and 73% of the VG preferred what they were assigned to (p = 0.3319). After simulation, 81% LG versus 67% VG preferred their assigned group (p = 0.0994).

DISCUSSION

In our study, we set out to determine if educational videos resulted in any difference in knowledge
application for medical students compared to traditional lecture in an emergency medicine undergraduate clerkship. We found no educationally significant differences in knowledge application, acquisition, and retention. Students were satisfied with the group they were assigned to, with no significant difference before or after the application test. Therefore, videos adhering to multimedia design principles could replace traditional lectures in producing similar knowledge acquisition and application in an OSCE setting.

Inconsistency in learning outcomes in multimedia learning literature could be partly due to inconsistent adherence to design guidelines, such as Mayer’s multimedia design principles based on cognitive load theory. Multimedia learning requires a substantial amount of cognitive processing in the verbal and visual channels. This cognitive processing is limited. Mayer’s guidelines decrease channel overload and increase cognitive processing capacity by removing extraneous and confusing material. In Mayer’s controlled experiments, students did better in tests when these individual principles were used, leading him to conclude that “cognitive load is a central consideration in the design of multimedia instruction.” This echoes previous study that suggests that sound pedagogy is more important than technology.

The strength of our study is that we adhered strictly to Mayer’s multimedia principles and attempted to keep curriculum confounders to a minimum. While a similar recent study with multimedia principle-based videos focused on satisfaction and attitudes, we used skill application in an OSCE setting as our primary outcome.

Our study contributed to the voices of previous work that videos adherent to multimedia design principles could replace traditional lectures. These videos could become reusable learning objects, freeing up valuable expert faculty time to afford interactive in-class learning time. Rather than comparing videos to lectures, future research should focus on what specific interactive elements would be suitable for in-class learning.

**LIMITATIONS**

There are a few limitations in our study. We did not randomize students individually, but used an alternating block sequence for the intervention (VG) and control group (LG) to allow for a contemporaneous control group with respect to clinical maturity and exposure to trauma cases. As the university assigned student groups before the study, it is possible that students with preferences to clinical topics might have requested to be in specific groups. This preference was not captured.

While half of VG and LG had exposure to trauma patients before our study through previous mandatory surgery rotations, we did not collect data on previous experience in assessing trauma patients. We also did not collect data on previous exposure to multimedia learning and simulation training.

These confounders could potentially affect our outcome, as those with previous exposure to trauma patients, simulation, or multimedia learning might have done better in the OSCE. As students were not blinded to the group they were assigned to, there is potentially bias in their performances.

The lecture (1.75 hours) was longer than the combined length of the videos (0.76 hours). While we attempted to keep the material identical (same slides, annotations, and script) and by auditing the first lecture, students asked questions during lectures. Those questions and subsequent answers therefore represented minor deviations from the planned script. Even though a tracker was used, we could not identify separate page visits to measure individual views.

The LG had a one-time exposure to the lecture and the VG could access the videos at any time. It is possible VG had accessed the videos multiple times, as some VG self-reported watching them repeatedly prior to the OSCE and posttest. This could have skewed the results toward VG via spaced repetition.

Cook has suggested that media-comparative studies such as this should not be done as there are confounders even if the content is identical. Instead, he suggested research is best done within rather than between levels of hierarchy (configuration, instructional method, and presentation) to answer the question of “how” and “when” to use e-learning effectively. Other authors lamented that the scope of evaluation was often limited to satisfaction or knowledge acquisition. Addressing the concerns of Cook about comparing different media, we attempted to keep potential confounders between the two methods (narrator, material, script, slides, vignettes, etc.) to a minimum.

**CONCLUSIONS**

Videos strictly adhering to Mayer’s multimedia design principles could replace traditional lectures in medical
student skill acquisition and application. Medical educators should use well-designed videos to transmit content and devote classroom time to active learning activities.

We thank the Clerkship Program and the Dean for Undergraduate Medical Program at the University of Ottawa and the University of Ottawa Simulation and Skills Centre for participation in this study.

References


Does Applicant Gender Have an Effect on Standardized Letters of Evaluation Obtained During Medical Student Emergency Medicine Rotations?

Jessica Andrusaitis, MD, MS\(^1\), Clelia Clark, MBA\(^1\), Soheil Saadat, MD, MPH, PhD\(^2\), John Billimek, PhD\(^1\), Sara Paradise, MD\(^2\), Alisa Wray, MD, MAEd\(^2\), Warren Wiechmann, MD, MBA\(^2\), Shannon Toohey, MD, MAEd\(^2\), and Megan Boysen-Osborn, MD, MHPE\(^{1,2}\)

ABSTRACT

Background: The standardized letter of evaluation (SLOE) in emergency medicine (EM) is one of the most important items in a student’s application to EM residency and replaces narrative letters of recommendation. The SLOE ranks students into quartile categories in comparison to their peers for overall performance during an EM clerkship and for their expected rank list position. Gender differences exist in several assessment methods in undergraduate and graduate medical education. No authors have recently studied whether there are differences in the global assessment of men and women on the SLOE.

Objectives: The objective of this study was to determine if there is an effect of student gender on the outcome of a SLOE.

Methods: This was a retrospective observational study examining SLOEs from applications to a large urban, academic EM residency program from 2015 to 2016. Composite scores (CSs), comparative rank scores (CRSs), and rank list position scores (RLPSs) on the SLOE were compared for female and male applicants using Mann-Whitney U-test.

Results: From a total 1,408 applications, 1,038 applicants met inclusion criteria (74%). We analyzed 2,092 SLOEs from these applications. Female applicants were found to have slightly lower and thus better CRSs, RLPSs, and CSs than men. The mean CRS for women was 2.27 and 2.45 for men (\(p < 0.001\)); RLPS for women was 2.32 and 2.52 for men (\(p < 0.001\)) and CS was 4.59 for women and 4.97 for men (\(p < 0.001\)).

Conclusions: Female applicants have somewhat better performance on the EM SLOE than their male counterparts.
In 1995, the Council of Residency Directors in Emergency Medicine (CORD-EM) established a task force to create a standardized letter of recommendation (SLOR). Renamed the standardized letter of evaluation (SLOE) in 2013, the intention of the SLOR/SLOE was to make a standardized evaluation that could stratify students into quantiles based on performance during an emergency medicine (EM) rotation.

The SLOE replaces traditional narrative letters of recommendation for applicants to EM residency programs. Program directors (PDs) regard the SLOE as the most important item used for selecting applicants to interview. Students typically include two or more SLOEs in their electronic residency service application (ERAS) to EM residency programs. While there are many benefits of the SLOE, it is unclear whether the SLOE can predict future performance.

The SLOE ranks students in various aspects of their EM clerkship performance (e.g., work ethic, teamwork, anticipated level of guidance needed). The global assessment section (section C) of the SLOE places the student into a quantile for two questions. Question C1 asks: “Compared to other EM residency candidates you have recommended in the last academic year the candidate is: top 10%, top third, middle third, lower third.” Question C2b asks: “How highly would you estimate the candidate will reside on your rank list: top 10%, top third, middle third, lower third, unlikely to be on our rank list.”

The SLOE global assessment scores may be inflated, with most students clustering in the higher categories. A study from 2004 found that female authors assessing female students tended to give higher global assessment scores on the SLOR than female authors assessing males or male authors assessing males or females. Currently, most SLOEs are written by a group of authors, including the clerkship and residency leadership, so the influence of author gender may be less relevant; however, a difference in SLOE scores for female and male students may still exist. Studies from undergraduate and graduate medical education have demonstrated gender differences associated with several assessment methods. For example, men in EM residencies achieve higher milestone assessments than women at graduation.

With the revised format of the SLOE and the tendency for more SLOEs to be completed by groups of EM residency and clerkship leaders, rather than individual faculty, we sought to determine if there is any difference in the SLOE global assessment of female versus male applicants. A recent study by Li and colleagues did not find any gender bias associated with the words used to evaluate female applicants; however, Li et al. did not study quantitative findings from SLOE rankings. Given the significance of the SLOE for applicants to EM residency programs, it is important to evaluate its objectivity, observing if any implicit bias exists among SLOE writers. As the primary outcome for our study, we sought to determine whether there was a difference in performance on SLOE global assessment scores between self-identified female and male applicants.

METHODS

We performed a retrospective review of SLOEs from applicants to the University of California at Irvine (UCI) EM residency program through the ERAS during the 2014–2015 and 2015–2016 application cycles. The UCI EM residency program is a 3-year program in an urban academic medical center in the greater Los Angeles metropolitan area.

We included records for applicants from all Liaison Committee for Medical Education Doctor of Medicine (MD) granting schools, excluding Puerto Rico and Canada. Two trained, nonblinded abstractors collected data using a standardized data abstraction form. The abstractors recorded Association of American Medical Colleges (AAMC) identification number, applicant gender, application year, medical school attended, United States Medical Licensing Examination (USMLE) step 1 score, step 2 clinical knowledge (CK) score, and data from each SLOE. From the SLOE, we collected the following: rotation location, date of rotation, author type (e.g., PD, clerkship director [CD], group, or faculty), number of letters authored in the previous year, comparative rank score (CRS), and rank list position score (RLPS). We did not record the gender of the letter writer, as most SLOEs were group SLOEs. The CRS and RLPS are the quantile answers (e.g., “top third”) to questions C1 and C2b, respectively, which are described in the introduction. The data abstractors and senior investigator held periodic meetings to discuss data abstraction and answer questions, which were resolved via consensus. The senior investigator sampled 5% of records to ensure quality. We calculated Cohen’s kappa for agreement. The senior investigator reviewed all final data for omissions and accuracy prior to analysis. The data were stored in a secure database, REDCap.
We screened and recorded data from SLOEs from all 4-week or 1-month traditional EM rotations (i.e., not pediatric EM or EM ultrasound). For the study analysis, we excluded subjects if they did not have any SLOEs in their applicant file. We excluded SLOEs if they were not written by an author who had written more than five SLOEs in the previous year, unless they predicted that they would write more than five in the current academic year. We excluded SLOEs that were not written by a faculty group, PD, CD, or any combination of these. We excluded SLOEs if they did not have complete data available.

For the CRS, we assigned a value to the following categories: top 10% (1 point), top third (2 points), middle third (3 points), and lower third (4 points). For the RLPS, we assigned the same values to each category and assigned 5 points to the “unlikely to be on our rank list” category. We calculated a composite score (CS) for each SLOE, which is the sum of the CRS and RLPS. The best CS possible is 2, which corresponds to “top 10%, top 10%,” whereas the worst possible CS is 9, which corresponds to “lower third, unlikely to be on our rank list.” We wanted to detect a difference of 15% in CSs for female and male applicants, which corresponds to an approximate single quantile difference (top third versus middle third) for a middle third, middle third candidate. Assuming a reliability of 50%, the required sample size was 167 participants in each group (334 total participants) for 80% power and significance of \( p = 0.05 \).

We describe the distribution of scores for the CRS and RLPS for female versus male applicants. We compared the mean CRS, RLPS, and CS for all SLOEs from students who identified as female versus those who identified as male, using the Mann-Whitney U-test. We used IBM SPSS Statistics for Windows, version 25.0, to analyze data. We obtained institutional review board (IRB) approval from UCI before the study commenced. The IRB did not require subject consent.

RESULTS

There were 4,066 total EM applicants from U.S. MD-granting schools in 2015 and 2016. The UCI Emergency Medicine Residency Program received 640 applications in 2015 and 768 applications in 2016, totaling 1,408 applicants. Of these, 1,053 were applicants from U.S. MD-granting schools. For the study analysis, we excluded 15 applicants because they did not have any SLOEs in their file that met inclusion criteria. We included 1,038 applications (74% of total applications) in the study analysis. All students reported either male or female gender. Most students included in the study identified as men \((n = 685, 66\%)\), which was similar to the percentage of male residents in EM in 2015 (63.1%). Agreement between the senior investigator and data abstractors was substantial \(( \kappa = 0.919–1.000 \))

Applicants had one \((n = 132, 13\%)\), two \((n = 644, 62\%)\), or three \((n = 262, 25\%)\) SLOEs in their files, totaling 2,206 SLOEs. We excluded 114 of these SLOEs because they had incomplete data; were not written by a group, PD, CD, or combination; or were not written by an author who had written more than

![Figure 1. Flowsheet for study inclusion and exclusion. CD = clerkship director; DO = doctor of osteopathic medicine; IMG = international medical graduate; PD = program director; SLOE = standardized letter of evaluation.](image-url)
five SLOEs in the previous year (Figure 1). Of the 2,092 included SLOEs, 1,386 (66%) SLOEs were from male applicants.

Students’ medical schools were in the following regions: west (n = 294, 28%), midwest (n = 236, 23%), northeast (n = 254, 24%), and south (n = 254, 24%). The mean (±SD) step 1 score for men in the study was 232.0 (±16.4) and was 228.2 (±16.7) for women (p < 0.001). The mean (±SD) USMLE step 2 score for men in the study was 241.8 (±14.6) and 243.7 (±13.9) for women (p = 0.035).

Women achieved slightly lower, and thus better, CRSs, RLPSs, and CSs than men. The mean CRS for women was 2.27 and was 2.45 for men (p < 0.001). The mean RLPS for women was 2.32 and was 2.52 for men (p < 0.001). The mean CS was 4.59 for women and 4.97 for men (p < 0.001). Nineteen percent of women received a “top 10%” rating on the CRS (n = 134) and the RLPS (n = 134); 15% of men received a top 10% CRS (n = 203) and 14% earned a top 10% RLPS (n = 195; p < 0.001). Figure 2 gives the distribution for each category rank in the SLOE by gender.

**DISCUSSION**

Differences in the assessment of men and women have been documented within undergraduate and graduate medical education. Most notably and recently, Dayal et al. studied the effect of gender on milestone assessments in EM residency training programs. Dayal and colleagues found that at the beginning of training, milestone assessments tend to be similar between men and women, with women achieving higher mean scores in many categories. At the end of training, however, men receive higher milestone assessments.

Our study was designed to observe if any gender differences exist at the earliest phase of EM training, the EM clerkship. We found that, from a large pool

![Figure 2. Distribution of SLOE CRS and RLPS for men and women. CRS = comparative rank score; RLPS = rank list position score; SLOE = standardized letter of evaluation](image)
of EM applicants, female students are assessed somewhat more positively than male students in the global assessment on the SLOE. Although our study findings were statistically significant, the difference between male and female CSs (−0.38) was relatively small and corresponds to one quantile difference (i.e., top third vs. middle third in one category) for every three men.

It is unclear what is driving the discrepancies between gender and performance at each stage of training. Women may perform slightly better during the clerkship (our findings). Both genders perform similarly during intern year on milestone assessments.20 Men may perform better during their final year of residency on milestone assessments.20 It is possible that qualities important for the success of a senior resident are different than the qualities important for success of a fourth-year medical student. It is possible that a person’s gender may offer qualities that predispose them to certain skills and attitudes. On the other hand, it is possible that one gender may be more negatively assessed than another gender when they exhibit certain qualities (e.g., assertiveness in women or uncertainty in men). Perhaps, in their role as subinterns, male medical students may suffer from gender bias and are perceived more negatively than their female counterparts. Alternatively, female medical students may perform better in this particular clinical environment. This could also be a function of different constructs being assessed in the SLOE versus milestones.

Haist et al.26 found that women performed better than men on clinical examinations and suggested it may be due to their improved ability to deal with uncertainty, which is much higher in clinical years than basic science years. While men had higher USMLE Step 1 scores, women had higher USMLE Step 2 CK scores in our study. This is consistent with previous studies on the USMLE and gender.28,29 The USMLE Step 2 CK score difference could be responsible for the gender difference observed in our study, since the USMLE Step 2 CK tests clinical knowledge and its application. Ultimately, the cause for our observed gender differences EM clerkship assessment is unclear.

Moving forward, it would be interesting to observe longitudinal assessments of medical students from admission to residency. As we move toward competency-based assessment for undergraduate and graduate medical education, it is essential that we acknowledge the risk of gender and other biases.

LIMITATIONS

We only included applications to a single residency program; however, the regional distribution of medical schools was diverse and there were a large number of applicants to our program. Our female study subjects performed better on the USMLE Step 2 CK. While this is consistent with national trends,28 our observed gender difference could be consistent with actual differences in clinical skills. We do not know if SLOE writers had USMLE Step 2 CK performance data available when assigning a RLPS. The leadership of our residency program is largely female and could have influenced the gender of applicants applying to our program. The gender distribution of residents during the study period was 52% female. Although our study findings were statistically significant, the difference between male and female CSs was less than one quantile difference.

CONCLUSIONS

The standardized letter of evaluation is an important part of a student’s application to emergency medicine residency. Women receive slightly lower, and thus better, mean rank list position score, comparative rank score, and composite score when compared to men. Women receive a top 10% and top third rating with higher frequency than men. It is unclear what accounts for this difference.

References


The Correlation Between Emergency Medicine Residents’ Grit and Achievement

Adriana Segura Olson, MD\textsuperscript{1,2}, Kelly Williamson, MD\textsuperscript{3}, Nicholas Hartman, MD, MPH\textsuperscript{4}, Navneet Cheema, MD\textsuperscript{1}, and Nathan Olson, MD, MAEd\textsuperscript{1,5}, for the Emergency Medicine Education Research Alliance (EMERA)

ABSTRACT

Background: Early identification of emergency medicine (EM) residents who struggle with educational attainment is difficult. In-training examination (ITE) scores predict success on the American Board of Emergency Medicine (ABEM) Qualifying Examination; however, results are not available until late in the academic year. The noncognitive trait “grit,” defined as “perseverance and passion for long-term goals,” predicts achievement in high school graduation rates, undergraduate GPA, and gross anatomy. Grit-S is a validated eight-question scale scored 1 to 5; the average of responses represents a person’s grit. Our objective was to determine the correlation between EM resident Grit-S scores and achievement, as measured by MCAT percentiles, ITE scores, and remediation rates.

Study Design and Methods: This was a 1-year prospective, multicenter trial involving ten EM residencies from 2017 to 2018. Subjects were PGY-1 to -4 EM residents. Grit-S scores, MCAT percentile, remediation rates, ITE scores, and the ITE score’s prediction of passing the ABEM Qualifying Examination were collected. Correlation coefficients were computed to assess the relationship between residents’ grit and achievement.

Results: A total 385 of 434 (88.7%) residents participated who completed the Grit-S as part of a larger study. The mean Grit-S score was 3.62. Grit positively correlated with the predicted likelihood of passing the ABEM Qualifying Examination ($r = 0.134, n = 382, p = 0.025$). There was no correlation between grit and remediation ($r = -0.04, n = 378, p = 0.46$) or grit and MCAT percentiles ($r = -0.08, n = 262, p = 0.22$).

Conclusions: The positive correlation between Grit-S scores and percent likelihood of passing the ABEM Qualifying Examination demonstrates grit’s potential to assist residency leadership in early identification of residents who may attain a lower ITE score.

INTRODUCTION

Early identification of emergency medicine (EM) residents who struggle with educational attainment can be difficult. The EM In-Training Examination (ITE) is administered every February. Residents who attain higher ITE scores have a higher likelihood of...
passing their American Board of Emergency Medicine (ABEM) Qualifying Examination.\textsuperscript{1,2} This information is incredibly valuable to program leadership as they seek to assess medical knowledge both for competency-based evaluations and to identify residents who may need additional guidance. Unfortunately, the results of the ITE are not available until late in the academic year resulting in a potential delay in identifying residents who may benefit from study plans and resources implemented at the beginning of residency to improve their medical knowledge.

In recent years, there has been increasing interest in using assessments of noncognitive traits to predict achievement.\textsuperscript{3–8} Noncognitive traits, including personality, have been shown to predict academic performance and educational achievement.\textsuperscript{3–8} The noncognitive trait “grit,” defined as “perseverance and passion for long-term goals,” is a means to quantify an aspect of personality.\textsuperscript{3} Beyond intelligence, grit entails the vigor, interest, and effort required to attain a goal through adversity and failure. It seeks to help answer the question of why individuals of equal intelligence may achieve different levels of success.\textsuperscript{3} Grit has been shown to be stable over time in the short-term in different populations,\textsuperscript{4,9} but does seem to increase over time in the long term.\textsuperscript{3} The Short Grit Scale (Grit-S) is a validated eight-question scale scored 1 to 5, with 5 being the highest score. The average of the responses represents a person’s grit; the higher the score, the “grittier” the individual.\textsuperscript{4}

In nonmedical domains, grit has been shown to predict educational attainment and success in multiple arenas including spelling bees, high school graduation rates, undergraduate GPA, and retention of recruits at West Point.\textsuperscript{3,4,10–13} However, studies evaluating grit in medicine are limited, especially with respect to achievement in graduate medical education. Grittier medical students have been found to have higher class ranks and to be more successful in gross anatomy and grit correlates with surgical residents’ psychological well-being.\textsuperscript{9,14–16} However, to the best of our knowledge, there has been little research evaluating grit’s correlation with academic achievement in graduate medical education, and no study to date has assessed the association of grit and achievement in EM residents. Because grit has been shown to correlate with achievement and academic success and less “gritty” individuals have lower achievement in other domains, we sought to evaluate whether EM residents’ grit would correlate with their medical knowledge and thus academic achievement. Specifically, the purpose of this study was to evaluate whether Grit-S scores correlated with residents’ ITE scores, MCAT percentiles, and need for remediation.

**MATERIALS AND METHODS**

**Study Design**

This study was a secondary analysis of a larger multicenter prospective educational trial investigating the effectiveness of a wellness didactic curriculum and involved 10 ACGME-accredited EM residency programs in the United States, conducted over 1 year from 2017 to 2018. The residency programs that participated in the study were geographically varied from the midwest, south, southeast, and west coast. The study was reviewed by each institution’s institutional review board and received approval at each site prior to study initiation. Eight sites were postgraduate year (PGY)-1 to -3 residencies and two sites were PGY-1 to -4 residencies.

**Subjects**

Eligible subjects for this study were PGY-1 to -4 EM residents at the participating programs during the study period. There were no exclusion criteria. Participation in the survey study was voluntary, and incentive for participation was nominal and site-specific.

**Study Protocol**

Written informed consent was obtained from all study participants. In February 2017 and February 2018, we administered the Grit-S (Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10399/full) as part of a larger survey that also included demographic information, the Maslach Burnout Inventory,\textsuperscript{17} the Prime-MD PHQ-2 depression screen, and a career satisfaction scale.\textsuperscript{18,19} Participants also self-reported their MCAT scores and year taken. Based on the Association of American Medical Colleges’ historical percentile ranks for the MCAT, self-reported raw MCAT scores were converted to percentiles for the corresponding year taken for uniformity.\textsuperscript{20} The survey was administered either as a paper survey or via online proprietary survey\textsuperscript{21} at the preference of each site investigator. Study principal investigators at each site collected data regarding participants’ ITE scores, their percent chance of passing the ABEM Qualifying Examination based on their ITE scores, and whether the resident underwent remediation, defined as “any extra work required for a resident above that required of all residents excluding...
counseling without additional work requirements.” ABEM’s reporting of ITE scores includes a raw score, percentile, and percent likelihood of passing the ABEM Qualifying Examination. Each resident was assigned a unique identifier known only to the individual participant and the respective site investigator. Follow-up for nonresponders was program-specific, either in person or via e-mail.

Data Analysis
Categorical outcomes were summarized with frequencies and percentages and continuously distributed outcomes were summarized with the sample size, mean, and standard deviation (SD). The assessment of grit stability over a 1-year period was assessed with a paired t-test. A Pearson’s product-moment correlation coefficient was computed to assess the relationship between residents’ Grit-S scores and their MCAT percentiles, the ITE score’s prediction of passing the ABEM Qualifying Examination, and the change in ITE score’s prediction of passing the ABEM Qualifying Examination over a 1-year period. A Spearman’s Rho correlation coefficient was computed to assess the relationship between residents’ Grit-S scores and their need for remediation. Multivariable ordinary least squares regression was used to model the effect of Grit-S scores, MCAT percentiles, and the need for remediation on the ITE score’s prediction of passing the ABEM Qualifying Examination.

RESULTS
A total of 385 of 434 residents completed the Grit-S as part of a larger study for a response rate of 88.7%. The mean (±SD) age of participants was 30 (±3.1) years. The participants were 69% male and 31% female (Table 1).

The mean (±SD) EM resident Grit-S score was 3.62 (±0.55; Table 1). Of the residents with at least one Grit-S score, 96 of 385 (24.9%) residents completed a Grit-S 1 year apart (2017 and 2018). The 289 residents that only had one Grit-S score did not complete two Grit-S scales either because they graduated before the second administration, they were not yet residents during the first administration, or because they did not participate in one administration of the survey. Grit-S remained stable from 2017 to 2018; there was no statistically significant difference between the Grit-S scores in 2017 (3.59 ± 0.46) and 2018 (3.58 ± 0.56; p = 0.8733; Table 2).

There was no correlation between Grit-S score and MCAT percentile (r = −0.077, p = 0.216) or Grit-S score and remediation (r = −0.050, p = 0.325). There was a positive correlation between Grit-S score and the ITE score prediction of passing the ABEM Qualifying Examination (r = 0.134, p = 0.025; Table 3).

When controlling for MCAT percentile and need for remediation, for each one-point increase in Grit-S score, there was an increase of 2.30 (95% confidence interval [CI] = 0.45 to 4.15) percentage points in the chance of passing the ABEM Qualifying Examination. When controlling for Grit-S score and need for remediation, for each 1% increase in MCAT percentile, there was an increase of 0.08 (95% CI = 0.01 to −0.015) percentage points in the chance of passing the ABEM Qualifying Examination. When controlling

Table 1
Demographics of Participants

<table>
<thead>
<tr>
<th>Participant Characteristics (n = 385)*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (±SD)†</td>
<td>30 (±3.1)</td>
</tr>
<tr>
<td>Gender, n (%‡)</td>
<td>Male 248 (69.1)</td>
</tr>
<tr>
<td></td>
<td>Female 111 (30.9)</td>
</tr>
<tr>
<td>Ethnicity, n (%)§</td>
<td>Caucasian 277 (74.7)</td>
</tr>
<tr>
<td></td>
<td>Mixed 25 (6.8)</td>
</tr>
<tr>
<td></td>
<td>Latino 18 (4.9)</td>
</tr>
<tr>
<td></td>
<td>East Asian 17 (4.6)</td>
</tr>
<tr>
<td></td>
<td>African American 13 (3.5)</td>
</tr>
<tr>
<td></td>
<td>South Asian 12 (3.2)</td>
</tr>
<tr>
<td></td>
<td>Middle Eastern 8 (2.2)</td>
</tr>
<tr>
<td></td>
<td>Pacific Islander 1 (0.3)</td>
</tr>
<tr>
<td>EM PGY, n (%)</td>
<td></td>
</tr>
<tr>
<td>PGY-1</td>
<td>167 (43.4)</td>
</tr>
<tr>
<td>PGY-2</td>
<td>109 (28.3)</td>
</tr>
<tr>
<td>PGY-3</td>
<td>91 (23.6)</td>
</tr>
<tr>
<td>PGY-4</td>
<td>18 (4.7)</td>
</tr>
<tr>
<td>Grit-S score, mean (±SD)</td>
<td>3.62 (±0.55)</td>
</tr>
</tbody>
</table>

PGY = postgraduate year.

*385 of 434 participants completed a Grit-S.
†Age at time of first completion of Grit-S, n = 385; 17 participants did not indicate age.
‡n = 359; 26 participants did not indicate sex.
§n = 371; 14 participants did not indicate ethnicity.
¶PGY at time of first completion of Grit-S.

Table 2
Assessment of Grit Stability Over 1-Year Period

<table>
<thead>
<tr>
<th>Grit</th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>February 2017</td>
<td>96</td>
<td>3.59</td>
<td>0.46</td>
<td>0.8733</td>
</tr>
<tr>
<td>February 2018</td>
<td>96</td>
<td>3.58</td>
<td>0.56</td>
<td>0.8733</td>
</tr>
</tbody>
</table>

*Paired t-test.
for Grit-S score and MCAT percentile, the need for remediation predicted a decrease of 9.78 (95% CI = –12.71 to –6.85) percentage points in the chance of passing the ABEM Qualifying Examination (Table 4).

There was no correlation between Grit-S score and change in ITE scores over a 1-year period, from 2017 to 2018 (r = –0.060, p = 4.27). Table 5 shows Grit-S scores and ITE percentiles by PGY.

**DISCUSSION**

In our study, we found a correlation between EM residents’ grit and the ITE score’s prediction of passing the ABEM Qualifying Examination. Since EM ITE scores have been shown to predict future success on the ABEM Qualifying Examination,¹ ² and predicting academic achievement in residents can be difficult,¹ ² ² ³ Grit-S scores administered early during training may be useful in helping to predict trainees’ success on the ABEM Qualifying Examination. Our findings corroborate previous studies demonstrating grit’s prediction of educational attainment and success in both nonmedical³ ⁴ ¹ ⁰ ¹³ and medical domains.⁹ ¹⁴ ¹⁶

While Grit-S scores may be helpful for medical educators and graduate medical education program leadership to identify residents early on who may perform well academically, our correlation between grit and the ITE’s prediction of success on the ABEM Qualifying Examination was weak to moderate (r = 0.134) and not as strong as we had anticipated. Furthermore, while Grit-S scores predicted an increase in the ITE’s prediction of success, the increase in percent chance per increase in Grit-S score was modest (2.30 percentage points). While likely multifactorial, there are likely other noncognitive traits not examined in this study that may have an influence on success on the ITE. For example, the noncognitive trait, self-control, defined as “the capacity to regulate attention, emotion, and behavior in the presence of temptation,”³⁴ while strongly correlated with grit, has been shown to be more important in the completion of shorter-term goals, rather than long-term achievement.³ ⁴ Residents likely view studying for the ITE as a short-term goal and therefore, self-control may have been another predictor of ITE success in addition to grit.

We selected to use the MCAT because it is unique from other medical examinations, including the ITE, in that it does not solely rely on recall of information but also tests analysis and reasoning skills.²⁵ In the initial study of grit by Duckworth et al.,³ researchers used SAT scores as a marker of general mental ability.²⁶ If we consider MCAT percentiles versus other medical tests to be a more similar marker of general mental ability, then the lack of correlation between grit

| Table 3 | Correlations Between Grit-S Score and Resident Achievement |
|---|---|---|
| | Correlation Coefficient | p-value |
| MCAT percentile* | –0.077† | 0.216 |
| Remediation | –0.050‡ | 0.325 |
| ITE Score prediction of passing ABEM Qualifying Examination† | 0.134† | 0.025 |

ABEM = American Board of Emergency Medicine; ITE = In-Training Examination.
* n = 253; 132 participants did not indicate MCAT score.
† Pearson’s correlation coefficient.
‡ Spearman’s rho.
† n = 279; 279 participants had both a 2017 Grit-S and a 2017 ITE score available.

| Table 4 | Ordinary Least-squares Regression of Predictors of ITE Score Prediction of Passing ABEM Qualifying Examination |
|---|---|---|
| | Coefficient | 95% CI | p-value* |
| Grit-S score | 2.30 | 0.45 to 4.15 | 0.015 |
| MCAT percentile | 0.08 | 0.010 to 0.15 | 0.033 |
| Remediation | –9.78 | –12.71 to –6.85 | 0.000 |

ABEM = American Board of Emergency Medicine; ITE = In-Training Examination.
*Regression analysis.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Grit-S Scores, ITE by PGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grit-S score, mean (±SD)*</td>
<td></td>
</tr>
<tr>
<td>PGY-1</td>
<td>3.67 (±0.56)</td>
</tr>
<tr>
<td>PGY-2</td>
<td>3.61 (±0.51)</td>
</tr>
<tr>
<td>PGY-3</td>
<td>3.55 (±0.57)</td>
</tr>
<tr>
<td>PGY-4</td>
<td>3.62 (±0.65)</td>
</tr>
<tr>
<td>ITE score, percentile (±SD)†</td>
<td></td>
</tr>
<tr>
<td>2017 ITE*</td>
<td>65.53 (±24.98)</td>
</tr>
<tr>
<td>PGY-1*</td>
<td>68.72 (±23.83)</td>
</tr>
<tr>
<td>PGY-2*</td>
<td>61.37 (±22.56)</td>
</tr>
<tr>
<td>PGY-3*</td>
<td>65.69 (±27.69)</td>
</tr>
<tr>
<td>PGY-4*</td>
<td>71.94 (±26.78)</td>
</tr>
<tr>
<td>2018 ITE**</td>
<td>62.49 (±26.39)</td>
</tr>
</tbody>
</table>

ITE = In-Training Examination; PGY = postgraduate year.
* n = 279; 279 participants had both a 2017 Grit-S and a 2017 ITE score available.
† n = 81.
‡ n = 93.
§ n = 87.
* n = 18.
** n = 176; 176 participants had both a 2017 Grit-S and a 2018 ITE score available.
and MCAT percentiles is consistent with the findings of Duckworth et al. findings that grit is not dependent on innate intelligence. In fact, grit was found to be negatively correlated with SAT scores, suggesting that more intelligent individuals might be slightly less gritty and therefore may not work as hard for achievement.

While not part of our primary objective, we found that while Grit-S scores did not correlate with the need for remediation, the need for remediation did predict success on the ITE, which would be expected as many residents undergo remediation for low ITE scores. Additionally, USMLE Step 2 CK (Clinical Knowledge) scores have been shown to predict passage of the ABEM Qualifying Examination on the first attempt. We did not collect data regarding residents’ USMLE scores in this study so it is possible that there may have been a correlation between grit and Step 2 CK scores that was not assessed in this study.

We also hypothesized that grittier residents might view a low ITE score as a setback and, therefore, persevere to increase their studying and thus be more likely to improve their ITE scores in 1 year compared to residents with lower grit. However, we found no correlation between grit and changes in ITE scores over 1 year. Grittier residents had initial higher ITE scores so they likely were already appropriately preparing and studying for the ITE and thus did not need as much improvement.

Our findings of the stability of EM residents’ Grit-S scores over a 1-year period are consistent with previous findings of stability in high school students and general surgery residents. This has important implications as the Grit-S score can be obtained early in training once to help predict success on the ITE throughout residency.

LIMITATIONS

There are important limitations to this study. While the multi-institutional design allowed for a relatively large sample size, it is still a convenience sample. There exists the possibility that residents may have inadvertently self-reported inaccurate MCAT scores as they were recalling scores from greater than 5 years before. Furthermore, there may have been the potential for self-bias in survey responses; it is possible that respondents may have filled out the survey in the way they thought they “should,” rather than how they truly felt. In addition, we found a lack of correlation between grit and rates of remediation. Our rate of remediation was 12.5%, which was much higher than the previously reported rate in EM of 4.4%. Because of the wide variation in defining remediation across EM programs, we used a broad definition for remediation. It is possible that if we had used a stricter, more narrow definition, we may have identified the residents who were at highest risk for struggling academically or professionally or may have identified residents needing further remediation. ITE scores are subject to numerous other influences that were not measured in this study including residents’ personal lives and clinical rotation schedule. Furthermore, residency programs have varying levels of curricula geared toward ITE preparation including review sessions and online question bank assignments that likely influenced residents’ ITE scores at each institution.

CONCLUSION

We identified a positive correlation between emergency medicine residents’ grit and the likelihood of passing the American Board of Emergency Medicine Qualifying Examination, as measured by the in-training examination. Although a weak to moderate correlation, the Grit-S scale may represent a valuable method for early prediction of academic success of emergency medicine residents and may be useful in identifying residents that may require close monitoring and coaching early in their postgraduate training. Our findings add to a deeper understanding of the complexity of resident achievement and also raise the possibility that other noncognitive traits may be worth exploring in the future for predicting success in residency.

The authors thank Stephen Hall for his support with the statistical analysis.

REFERENCES


Supporting Information
The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10399/full
Data Supplement S1. Short Grit Scale.
Salivary Cortisol Concentrations, Grit, and the Effect of Time

Matthew L. Wong, MD, MPH,1 Gregory Peters,1 Joshua W. Joseph, MD, MS,1 Arlene Chung, MD, MACM,2 and Leon D. Sanchez, MD, MPH1

ABSTRACT

Background: Stress is a common experience in the emergency department (ED) and is a balance of personal capabilities versus demands. Residency training is meant to improve individuals’ capabilities and therefore may attenuate an individual’s stress response. Grit is a personality trait that may attenuate stress in individuals. In this study we explore the relationship between time of year, postshift salivary cortisol concentrations, and the influence of grit among attendings and residents in an academic ED.

Methods: Thirty-nine residents and 17 attendings were enrolled and followed for two academic years at an urban academic medical center. Postwork salivary cortisol samples were collected quarterly, and Duckworth 12-point Grit Scales were administered annually. Data from the relative quarters of the first and second years were combined, and the results were analyzed using descriptive statistics and generalized estimating equations.

Results: We analyzed 216 cortisol samples and 87 grit scores over 2 academic years. Between the first and fourth quarters of the academic year, the percentage of subjects with a detectable postshift cortisol sample decreased from 47.6% (30 of 62) to 18.4% (9 of 49). In the fourth academic quarter compared to the first, the odds that an individual had a detectable cortisol were significantly lower overall (odds ratio [OR] = 0.25, 95% confidence interval [CI] = 0.10 to 0.59, p = 0.01), including the subset of only trainees (OR = 0.26, 95% CI = 0.07 to 0.88, p = 0.01), and adjusting for grit did not meaningfully change the ORs.

Conclusion: Over the course of the academic year, the odds that an individual’s postshift salivary cortisol concentration will be above detectable concentration significantly decrease over time, and this relationship is not confounded by grit.

Although stress is universal to the practice of emergency medicine, the exact stressors for each individual are highly context specific and subjective. Stressors vary depending on the person, place, time, and event. Stress arises when an individual is confronted with a situation requiring more personal resources than they have available.1,2 All psychological stressors cause activation of the central nervous system, which stimulates the adrenal glands to secrete cortisol and catecholamines.3–6 These hormones have pleiotropic physiological and psychological effects that are sometimes contradictory. For example, there can be improvement in mental focus, but also decrease in working memory and cognitive performance.6–8

The benefit of supervised clinical experience in graduate medical education is that it safely increases residents’ abilities over time, even in the face of rising demands. The exact same stressors that overwhelm an
The intern may not even be consciously perceived by a senior resident. Because of the close relationship between stress and cortisol, a change in response to stress may be able to be detected as a change in cortisol concentrations over time. Some people are likely naturally more adept at managing stress than others. Grit is a noncognitive personality trait that describes the ability to prioritize long-term goals and may also mitigate an individual’s perception of stress. In this study we hypothesized that participants’ cortisol concentration will decrease over time as they gain more experience, training, and familiarity with emergency medicine. We also hypothesized that grit would predict each individual’s response to stress. We report the results of following residents and attending physicians at an urban academic emergency department (ED) over 2 years.

**METHODS**

**Study Setting and Population**

Subjects were recruited from an academic ED in June 2016 and June 2017 at the beginning of the academic year. The principal investigator (PI) presented the study’s design and goals in a 10-minute presentation at the beginning of the academic year during the scheduled weekly didactic time, and three additional recruitment e-mails were sent to all residents and attendings. All residents and attendings were solicited for enrollment during these presentations and e-mails. Interested subjects were referred to a trained research assistant to perform the study’s written consent, to avoid the PI or any co-investigator from directly consenting a trainee or colleague. The only exclusion criteria was chronic exogenous glucocorticoid administration. None of the subjects who referred themselves to the research team were excluded. There was no financial compensation for participation.

**Study Protocol**

The study was conducted at a tertiary academic hospital in an urban setting that takes care of approximately 56,000 patients per year. All patients are seen by residents primarily who then present the patient to an attending.

Participants privately answered the Duckworth 12-item Grit Scale in the beginning of the academic year and once per annum. Salivary cortisol measurements were collected from participants with the assistance of a trained research assistant every quarter of the academic year. Saliva samples were taken between 23:00 and 00:00 after working an evening shift at the host institution. All shifts were either from 14:00 to 23:00 or from 15:00 to 23:00. Participants were instructed to abstain from eating, drinking, or brushing their teeth 30 minutes prior to specimen collection. Participants were given a commercial Sarstedt Salivette container, which contains an absorbent cotton swab. Salivary cortisol analysis was performed by the pathology department using an electrochemiluminescence immunoassay in a Cobas c6000 analyzer (Roche Diagnostics).

Data were entered in a secure REDCap database administered by our institution on secure servers. Only the PI had access to individual subject-level data. Data analysis were performed with Stata 9.0.

**Data Analysis**

The respective quarters from each of the 2 years were combined and were analyzed in aggregate (i.e., chronological quarters 1 and 5, 2 and 6, 3 and 7, and 4 and 8 were combined). Grit scores were dichotomized into high versus low scores relative to the mean grit score of the entire study population. To determine the odds of a positive salivary cortisol sample, we performed a regression using generalized estimating equations using a logit link, a binomial distribution, and an exchangeable correlation structure. Generalized estimating equations are useful when examining data that was repeatedly collected from the same individuals, and we have used it in similar analyses. Additional analyses were performed to explore the relationship between detectable cortisol concentration, time of year, and the influence of grit on the relationship. The primary analyses of interest used data from all study subjects. A secondary analysis was performed using only the restricted subset of trainees only.

**Institutional Review**

This study was approved by the institution’s Committee on Clinical Investigations and the Department of Emergency Medicine’s Medical Education Executive Committee.

**RESULTS**

We enrolled 39 residents and 17 attendings between the 2 years. All subjects enrolled during the first year reconsented to participate the second year, and new
subjects were also enrolled the second year. All subjects enrolled during the first year were followed for a full eight quarters, and subjects enrolled the second year were followed for four quarters, for a mean (±SD) of 6.7 (±1.86) quarters. No subjects dropped out after consenting to participation. Research assistants contacted subjects every quarter to attempt to coordinate data collection, but not all 56 subjects were available during all quarters of the study period. Participants completed a total of 87 Grit Scales with a mean score of 3.7 (95% confidence interval [CI] = 3.6 to 3.8). Twenty-five (28.7%) of the Grit Scales were completed by PGY-1s, 21 (24.1%) were completed by PGY-2s, 11 (12.6%) were completed by PGY-3s, and 30 (34.5%) were completed by attendings (Table 1).

Of 420 possible subject-quarters, a total of 313 saliva samples were distributed to subjects. Of those, 87 samples were not returned, and 10 samples were collected but had an insufficient volume of saliva to analyze and were treated as missing. A total of 216 samples were available for final analysis. Fifty-six (25.9%) samples were taken from PGY-1s, 44 (20.4%) from PGY-2s, 29 (13.4%) from PGY-3s, and 87 (40.3%) from attendings. Mean concentration of the samples was 0.371 μg/dL (0.019–13.12 μg/dL; SD ± 1.44 μg/dL).

Due to the high incidence of undetectable cortisol concentrations, we dichotomized the samples into undetectable and detectable groups for analysis. A total of 135 saliva samples had undetectable cortisol concentrations and 81 had detectable cortisol concentrations. The mean (±SD) age of subjects at the time of the saliva collection was 31.0 (±5.6) years old. We collected 63 (29.2%) cortisol samples during the combined first academic quarters of the 2 years (i.e., chronologically quarters 1 and 5), 51 (23.6%) during the combined second quarters, 53 (24.5%) during the combined third quarters, and 49 (22.7%) during the combined fourth quarters of the year. Samples that were quantity not sufficient were distributed as such: one in the first chronological quarter, three in the second quarter, two in the third quarter, one in the fifth quarter, one in the seventh quarter, and two in the eighth quarter.

The proportion of undetectable cortisol samples increased from 52.4% (33/63) to 81.6% (40/49) over time from the combined first quarters to fourth quarters (Table 2). The odds of having a detectable cortisol concentration in the fourth quarters relative to the first were significantly decreased (odds ratio [OR] = 0.25, 95% CI = 0.10 to 0.59, p = 0.01) even after adjusting for the effect of high-grit status (OR = 0.23, 95% CI = 0.10 to 0.56, p = 0.01). Inclusion of an interaction term between high-grit status and time into the model was not statistically significant (p = 0.818). A box plot of subjects’ Grit Scales stratified by detectable versus undetectable cortisol concentrations by quarter is summarized in Figure 1. Post hoc one-way power analysis confirmed a sufficiently large sample size to detect a change in the proportion of negative cortisol values (β = 0.93).

In our secondary analysis of residents only, the odds of having a detectable cortisol concentration in the fourth quarters relative to the first were significantly decreased, without and with adjustment for grit (OR = 0.26, 95% CI = 0.07 to 0.88, p = 0.03; OR = 0.24, 95% CI = 0.07 to 0.85, p = 0.03, respectively). Inclusion of an interaction term between high-grit status and time into the model was not statistically significant (p = 0.98). These results are summarized in Table 3.

### DISCUSSION

In this study we serially measured postshift salivary cortisol concentrations of resident and attending emergency physicians and found that the odds of detecting cortisol decreased over time. This relationship was not confounded or modified by grit, and a similar trend was observed in both residents and attendings. These results provide surrogate evidence that stress in individuals decreases over the course over the year in an academic ED. This study contributes to the literature

| Table 1
| Descriptive Characteristics of Study Population Accounting For Multiyear Enrollment of Subjects |
|---------------------------------|-----------------------------------|
| Number of unique subjects       | 56                                |
| Number of unique trainees       | 39                                |
| Subject age at sample collection, mean (±SD) | 31.0 (±56.0) |
| Subject-quarters of data       | 313                               |
| Cortisol samples, total         | 216                               |
| PGY-1                           | 56 of 216 (25.9%)                 |
| PGY-2                           | 44 of 216 (20.4%)                 |
| PGY-3                           | 29 of 216 (13.4%)                 |
| Attending                       | 87 of 216 (40.3%)                 |
| Grit scores, total (mean, 95%CI) | 87 (3.7, 3.6–3.8) |
| PGY-1                           | 25 of 87 (28.7%)                  |
| PGY-2                           | 21 of 87 (24.1%)                  |
| PGY-3                           | 11 of 87 (12.6%)                  |
| Attending                       | 30 of 87 (34.5%)                  |

Wong et al. • SALIVARY CORTISOL CONCENTRATIONS, GRIT, AND EFFECT OF TIME
by examining physiologically quantifiable measurements of stress.

Stress is ubiquitous in the practice of emergency medicine. Although the most salient sources of stress may vary from one emergency physician to another, the experience of stress in emergency medicine is universal. Stress is not always a bad thing; some amount of stress is useful because it promotes arousal and helps concentration on tasks. However, when left unchecked, stress has many negative cognitive, emotional, and physical consequences. The epidemic of burnout in emergency medicine has highlighted the unhealthy balance of beneficial and detrimental psychological forces affecting physicians, such as stress.

Discussions about stress, burnout, and mental health in medical professionals are more common now than ever before the literature, the medical community, and the lay press. Emergency physicians are some of the most at-risk physicians in medicine, and this topic is pertinent to practicing physicians, physicians in training, and those who are responsible for training and leading physicians. The trainees in our study had fewer samples with detectable cortisol at the end of the year compared to the beginning. This is an interesting finding because salivary cortisol measurements have never been used in this way, in a population like this. Furthermore this phenomenon may reflect an improvement in their general wellness, or it may be a reflection of their training. Future research into wellness, stress inoculation training, and other performance-enhancing psychological skills may benefit from an accurate and objective means of an individual’s, as well as a group’s, level of stress. Developing the capacity to handle stressful medical and traumatic resuscitations is a hallmark of

<p>| Table 2 |
| Salivary Cortisol Samples Above and Below Detectable Assay Threshold by Quarter, Stratified by Training Status |</p>
<table>
<thead>
<tr>
<th>Salivary Cortisol Concentration</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below assay</td>
<td>33 (52.4%)</td>
<td>25 (49.0%)</td>
<td>37 (69.8%)</td>
<td>40 (81.6%)</td>
<td>135 (62.5%)</td>
</tr>
<tr>
<td>Above assay</td>
<td>30 (47.6%)</td>
<td>26 (51.0%)</td>
<td>16 (30.2%)</td>
<td>9 (18.4%)</td>
<td>81 (37.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
<td>51</td>
<td>53</td>
<td>49</td>
<td>216</td>
</tr>
<tr>
<td>Trainees only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below assay</td>
<td>24 (58.5%)</td>
<td>14 (46.7%)</td>
<td>25 (78.1%)</td>
<td>22 (84.6%)</td>
<td>85 (65.9%)</td>
</tr>
<tr>
<td>Above assay</td>
<td>17 (41.5%)</td>
<td>16 (53.3%)</td>
<td>7 (21.9%)</td>
<td>4 (15.4%)</td>
<td>44 (34.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>30</td>
<td>32</td>
<td>26</td>
<td>129</td>
</tr>
</tbody>
</table>

| Figure 1. Grit scale versus detectable (+) or undetectable (–) post-shift salivary cortisol level, for each combined academic quarter (Q1–Q4). |

Table 3
Odds of a Detectable Cortisol Sample Above Assay by Quarter, With and Without Adjustment for Grit

<table>
<thead>
<tr>
<th></th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All subjects (n = 232)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted model</td>
<td>Reference</td>
<td>1.13 (0.54–2.35)</td>
<td>0.47 (0.22–1.0)</td>
<td>0.25 (0.10–0.59)</td>
</tr>
<tr>
<td>Adjusted model</td>
<td>Reference</td>
<td>1.1 (0.50–2.23)</td>
<td>0.45 (0.21–0.97)</td>
<td>0.23 (0.10–0.56)</td>
</tr>
<tr>
<td>Trainees only (n = 134)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted model</td>
<td>Reference</td>
<td>1.62 (0.63–4.18)</td>
<td>0.40 (0.14–1.12)</td>
<td>0.26 (0.07–0.88)</td>
</tr>
<tr>
<td>Adjusted model</td>
<td>Reference</td>
<td>1.53 (0.59–4.02)</td>
<td>0.38 (0.13–1.08)</td>
<td>0.24 (0.07–0.85)</td>
</tr>
</tbody>
</table>

Data are reported as OR (95% CI).  *p = 0.01;  **p = 0.01;  ***p = 0.03;  ****p = 0.03.
emergency medicine, and salivary cortisol measurements may be a means to measure that ability.

In our study, individuals with all levels of grit exhibited the same trend of decreasing salivary cortisol concentrations over time. This is interesting because we suspected that individuals with little grit might have had a different trajectory than those with lot of grit. The concept of grit has become very popular in the lay press and may be a useful concept in cognitive performance, but it is still an area of active investigation. Some research has shown that grit in emergency physicians is predictive of burnout. Grit may be a useful measure to assess risk of prematurely departing residency training early. Other work has shown that levels of grit in emergency physician are about the same as those in other medical and surgical specialties and West Point cadets.

These results are important because if salivary cortisol concentrations were found to be reliable and predictive of stress, then cortisol could be used as a metric in education and training, as well as physician well-being. This study also provides some additional findings about the importance of grit in emergency physicians. More work is needed to establish if quantitative physiological or psychological measures of stress are valid or useful.

LIMITATION

This study is limited by the relatively small sample size, the fact that it was conducted at a single site, and the nonrandom self-selection of study participants. There is no clear consensus on when or how cortisol measurement should be used in measuring stress over an entire shift, and postshift salivary cortisol samples may not be the most sensitive or specific measurement. We did not control for recent shift burden or recent shift type when coordinating which dates to collect saliva from the participants. Potential dates were coordinated between the research assistant and the subject to find a mutually agreeable time, but it was not specifically required that it was the first shift of its type in a set of shifts or any other arrangement. The effect of proximal shift burden (or lack of burden) and its effect on subsequent salivary cortisol concentrations is unknown, and given the volume of shifts that residents and attendings work, and the permutations of shift order, it would have been logistically impossible to schedule seemingly commensurate dates among subjects.

The concept of grit has been critiqued and may not provide any more additive information about a person’s personality than can be gleaned by other means. Furthermore, a holistic framework encompassing stress, cortisol, well-being, and cognitive performance has not been clearly defined.

Future multicenter work involving more trainees would be needed to validate these findings. Additionally, although we restricted the sample collection to similar shifts at a single institution, it was not possible to ensure that all of the shifts for the subjects were all equally stressful. Further adjustment for on-shift stressors would provide more insight, but it begs the question of this current project and how to measure stress quantitatively.

CONCLUSIONS

Detectable concentrations of salivary cortisol are common in the beginning of the academic year, but the incidence decreases over time, suggesting an effect of training. Individuals with low grit and individuals with high grit share this same trend. This effect is observed in both trainees and attendings, and further investigation is needed to examine the role of stress and resilience in emergency medicine trainees.

References

The Equivalence of Video Self-review Versus Debriefing After Simulation: Can Faculty Resources Be Reallocated?

Gregory J. Tudor, MD, FACEP1,2, Gregory S. Podolej, MD1,2,3, Ann Willemsen-Dunlap, PhD2,3, Vivian Lau, MD1,2, Jessica D. Svendsen, CCRC3, Jeremy McGarvey, MS3, John A. Vozenilek, MD, FACEP1,3, and Lisa T. Barker, MD, FACEP1,3

ABSTRACT

Introduction: Traditional simulation debriefing is both time- and resource-intensive. Shifting the degree of primary learning responsibility from the faculty to the learner through self-guided learning has received greater attention as a means of reducing this resource intensity. The aim of the study was to determine if video-assisted self-debriefing, as a form of self-guided learning, would have equivalent learning outcomes compared to standard debriefing.

Methods: This randomized cohort study consisting of 49 PGY-1 to -3 emergency medicine residents compared performance after video self-assessment utilizing an observer checklist versus standard debriefing for simulated emergency department procedural sedation (EDPS). The primary outcome measure was performance on the second EDPS scenario.

Results: Independent-samples t-test found that both control (standard debrief) and intervention (video self-assessment) groups demonstrated significantly increased scores on Scenario 2 (standard—t(40) = 2.20, p < 0.05; video—t(45) = 3.88, p < 0.05). There was a large and significant positive correlation between faculty and resident self-evaluation (r = 0.70, p < 0.05). There was no significant difference between faculty and residents self-assessment mean scores (t(24) = 1.90, p = 0.07).

Conclusions: Residents receiving feedback on their performance via video-assisted self-debriefing improved their performance in simulated EDPS to the same degree as with standard faculty debriefing. Video-assisted self-debriefing is a promising avenue for leveraging the benefits of simulation-based training with reduced resource requirements.

Procedural sedation is a core competency for the practice of emergency medicine comprising a specific competency milestone in the Accreditation for Graduate Medical Education Next Accreditation System.1 Despite advances in technology such as end-tidal CO2 monitoring, the safety profiles of commonly used...
medications, and the implementation of published practice guidelines, adverse respiratory events are still observed.2–4

Immersive simulation provides a safe training environment for learning procedural sedation. The traditional learning model for simulation-based training includes a postsimulation debriefing in which learners’ experiences are explored and reflected upon with the goal of improving future performance.5–7 Such a format is resource-intensive due to the higher faculty–trainee ratio when compared to other educational strategies, and there is increasing pressure to evaluate and report the costs associated with simulation-based educational interventions.8

Shifting the degree of primary learning responsibility from the faculty to the learner through self-directed learning has received greater attention as a means of reducing resource intensity.9 While the benefits of self-directed learning, the optimal uses of simulation for self-directed learning in clinical education in this manner has not yet been fully explored.10–12 The purpose of the study was to determine if video-assisted self-debriefing, as a form of self-directed learning, would have equivalent learning outcomes compared to standard debriefing.

The Society for Simulation in Healthcare defines debriefing as a formal, collaborative, reflective process within the simulation learning activity that encourages participants’ reflective thinking and provides feedback about their performance while various aspects of the completed simulation are discussed.13 We propose that self-reflective video “debriefing” is a form of debriefing because it meets the criteria of the above definition. Specifically, it is formal in that it is a structured experience, it is collaborative in that the checklist is a derivation of a modified Delphi panel, and it is reflective in that viewing one’s own performance on video with a checklist provides a framework for comparison within the simulation learning activity.

**METHODS**

**Study Design**

This randomized cohort study consisting of 49 PGY-1 to -3 emergency medicine residents compared performance after video self-assessment utilizing an observer checklist versus standard debriefing for simulated emergency department procedural sedation (EDPS). The study was approved by the institutional review board under expedited review criteria. Data for research were only extracted from participants who voluntarily provided consent.

The study consisted of 49 PGY-1 to -3 emergency medicine residents. Participation in the residency program’s formal biannual training in procedural sedation was included in demographic data. Eligible study subjects (all emergency medicine residents) were randomized to intervention or control groups prior to enrollment. Learners participated in individual simulations that were deployed as part of weekly residency conference. Residents were approached by designated research personnel prior to participation to obtain consent to include their deidentified data.

An observer checklist (Data Supplement S1, Appendix S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10372/full) was created that assessed clinical, pharmacologic, and communication areas of EDPS with a total of 49 checklist items. The checklist was developed using a two-round modified Delphi process.14 The checklist was piloted with PGY-3 residents and revisions were made to increase reliability. Inter-rater reliability was determined with five EDPS videos using the Brennan-Prediger kappa variant, and the overall mean kappa for all questions and subtypes was 0.90.

The “Quality of Patient Care Assessment” utilized a five-point Likert scale. This was a global assessment performed by both individual learners and two emergency medicine core faculty members.

Video-assisted self-debriefing was structured as a self-guided postevent debriefing12 in which learners viewed video recording of their performance while completing the observer checklist. Standard debriefing was defined as postevent facilitator-guided,12 using the “debriefing with good judgment” framework.15 All debriefers (GJT, LTB) were emergency medicine physicians trained in debriefing. Debriefers completed the observer checklist during the simulation to use as a debriefing guide.

**Scenario Design**

Two variants of an EDPS scenario were created for the study. Scenario 1 involved a 6-year-old child with a dog bite to the thigh that required suturing. Scenario 2 involved an 18-month-old child with an abscess of the inner thigh that required incision and drainage.

Both variants had the same objectives:
1. Perform standard elements of pre-sedation evaluation for a pediatric patient requiring procedural sedation and analgesia (PSA);
2. Correctly identify and correct medication error embedded in scenario;
3. Correctly identify and correct hypoventilation from oversedation embedded in scenario;
4. Provide standard-of-care sedation (preprocedural, procedural, and postprocedural) for a child with a condition requiring PSA (see checklist).

To increase fidelity and allow the test subjects to focus solely on the procedural sedation, three trained standardized participants (SPs) were utilized in each case. Their specific training was uniform and the scripted dialogue is included in Data Supplement S1, Appendix S2. There were no more than three different SPs in the role of the parent, limiting variation in performance. In addition, each of the SP parents had an earpiece to provide any necessary correction during the simulation. These SPs remained the same throughout the study.

Protocol
All eligible subjects participated in the simulated EDPS. The intervention group (n = 25) used video self-assessment while the control group (n = 24) underwent standard debriefing. Randomization assignment was revealed directly after participation in the simulation so that raters were blinded until the simulation scenario was complete. For the video group, research personnel provided acclimation to the video playback features and the observer checklist. The same performance checklists were used for self-guided learning and in faculty debriefing to ensure the same content was covered.

The primary outcome measure was performance in the second EDPS scenario. Case order was the same for both groups. All participants ran Scenario 2 (incision and drainage) within a period of 4 months from their participation in Scenario 1. Both groups experienced facilitator-led debriefing for Scenario 2. Change in checklist scores between the two simulations for each group was calculated and compared using paired and independent-samples t-test. Statistical analysis was performed with R software (The R Foundation).

RESULTS
Demographic information for the control and intervention groups is shown in Table 1. The lack of a significant score difference between randomizations for Scenario 1 indicates that they started at the same baseline.

The means and standard deviations (SD) for the faculty evaluations of the standard and video randomizations for Scenario 2 (incision and drainage) can be seen in Table 2 below. Independent-samples t-test found that both groups demonstrated significantly increased scores on Scenario 2 (video—t(45) = 3.88, p < 0.05; standard—t(40) = 2.20, p < 0.05).

A linear regression was used to determine the effect of randomization, scenario, and the interaction between the two on performance score. No significant associations were found between score and randomization or between randomization and scenario. These results indicate that the observed increase in performance on the second scenario (laceration repair) compared to the first scenario (incision and drainage) was independent of randomization. The conditional plots in Figure 1 graphically display the relationship

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographics for Control and Intervention Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video Self-assessment Group (n = 25)</td>
</tr>
<tr>
<td>PGY level</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22 (88.0)</td>
</tr>
<tr>
<td>2</td>
<td>2 (8.0)</td>
</tr>
<tr>
<td>3</td>
<td>1 (4.0)</td>
</tr>
<tr>
<td>Number of days from sedation course to Scenario 1</td>
<td>394.04 (245.58)</td>
</tr>
<tr>
<td>Scenario 1 score</td>
<td>55.88 (±16.33)</td>
</tr>
<tr>
<td>Number of days between Scenario 1 and Scenario 2</td>
<td>79.0 (±51.9)</td>
</tr>
</tbody>
</table>

Data are reported as n (%) or mean (±SD).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Postintervention Performance Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video Self-debriefing</td>
</tr>
<tr>
<td>Scenario</td>
<td>n</td>
</tr>
<tr>
<td>Video self-debriefing</td>
<td></td>
</tr>
<tr>
<td>Lac</td>
<td>25</td>
</tr>
<tr>
<td>I&amp;D</td>
<td>22</td>
</tr>
<tr>
<td>Standard debriefing</td>
<td></td>
</tr>
<tr>
<td>Lac</td>
<td>24</td>
</tr>
<tr>
<td>I&amp;D</td>
<td>18</td>
</tr>
</tbody>
</table>

I&D = incision and drainage; Lac = laceration repair.
There was a baseline difference between the video and standard debriefing group for the laceration repair. The standard debriefing group started out with a better performance (though not statistically significant). Even though the video group started out five points below the baseline, they demonstrated clear improvement (there was no overlap in error bars between Scenario 1 and Scenario 2).

**Paired Resident Self-evaluation Versus Faculty Evaluations**

Chi-square and Fisher’s exact tests were used to determine if there was a difference in the frequency at which resident self-evaluation and faculty assessments identified specific items on the checklist as “performed.” Of the 21 performed checklist items that had both faculty and resident evaluations, a significant difference between resident self-evaluation and faculty evaluations was found only for the “jaw thrust performed” item with 72% of residents classifying it as performed in contrast to only 40% of faculty ($\chi^2(1, N = 50) = 5.19, p < 0.05$).

Total checklist scores were calculated by assigning one point for each completed action. Team performance questions were excluded from this analysis. There was no significant difference between faculty and residents self-assessment mean scores ($t(24) = 1.90, p = 0.07$).

**DISCUSSION**

Kolb described the use of experiential learning in the “critical linkages among education, work, and personal development. His study of the similarities of several learning theories emphasized the linkage of experience to reflection, which in turn allows the learner to “[adapt] to one’s total life situation.”16 In this study, residents are prompted to engage in a deeper and guided self-reflection assisted by video. From a pedagogic standpoint, this self-reflection is a transforming experience leading to “new implications for action.” This study addresses, in part, the exploration of a resource-sparing reflective process which retains the benefit of the experience for the learner.

Video-assisted self-debriefing may be one strategy within a larger context of self-regulated learning (SRL). A key feature for success in SRL is deliberate planning of activities to support individual progress rather than merely “learning alone.”17 In our study, both scenarios lent themselves to the use of a performance checklist for assessment. Similarly, proper scaffolding of self-guided mastery learning has demonstrated educational outcomes equivalent to instructor led mastery-learning interventions for ACLS training of internal medicine residents with lower associated costs.18

The immediate appeal of self-directed learning is the logistic flexibility and reduction of costs. A recent study calculated significant cost savings for video-assisted self-debriefing when compared to instructor-led debriefings of anesthesiology residents participating in “perioperative crisis scenarios.”18 While there is cost associated with purchasing and utilizing video recording equipment, simulation recording is a common practice and part of the baseline equipment at our institution. Furthermore, cost-effectiveness must take into consideration the intangible expenses associated with faculty time commitment, faculty scheduling burden, and training faculty to become skilled in simulation debriefing. In addition, there is a significant time investment required to develop, deploy, and study a
checklist-based educational intervention. As a checklist is developed it may be validated then utilized across multiple learner groups with similar educational goals (e.g., residents in emergency medicine, pediatrics, surgery).

Video-assisted debriefing can be accomplished in almost any environment—the learner sits at a laptop with headphones and views his or her recording with the checklist. Cost savings for faculty debriefing time could be a significant benefit for appropriately selected simulation scenarios amenable to video checklist debriefing. In addition, time saved by video debriefing may be spent on learner remediation, scenario refinement, or other educational activities. In our study, each scenario had an average length of 14 minutes and 18 seconds per participant.

While these results demonstrate positive impact of video-assisted self-briefing on learners’ subsequent performance, video review in and of itself is not the panacea. Martin et al.19 report discordant ability to assess self versus others performance of a simulated patient communication encounter via video. However, the assessment tool supporting video review in that study was a behavioral rating scale, which allows for greater rating subjectivity. Our data demonstrate that for procedural sedation, video checklist debriefing can result in self-assessments highly correlated with faculty assessments and, more importantly, similar improvement in performance when compared with standard debriefing. Our results suggest that video self-review may be more beneficial for novice (lower score) learners. Expert learners in the standard debriefing group started higher and had a smaller improvement which suggests that expertise is asymptotic (it takes proportionally much more effort to get from 90 to 100 than it does from 10 to 100) and that expert learners may benefit from a more nuanced faculty guided debriefing.

In addition, our data demonstrate that for procedural sedation, video checklist debriefing can result in self-assessments that correlate highly with faculty assessments. More importantly, this equates to a similar improvement in performance when compared with standard debriefing. Our study may have been able to detect the improvement between scenarios because a checklist was used instead of a behavior rating scale.

LIMITATIONS

Several limitations are present. There were six individuals who were lost to follow-up in the standard debriefing group. Despite the significant positive impact on subsequent performance, the inclusion of these subjects may have potentially changed the results for the standard debriefing group relative to the intervention group. Our scenarios involved only two cases with PSA as the substrate for clinical care. The singular focus of PSA and the ability to create a binary checklist may not be generalizable to other case scenario types. Additionally, both case scenarios involved PSA for simple procedures (laceration repair and

![Figure 2. Comparison of resident vs faculty self-evaluation.](image-url)
abscess drainage), which may not represent the cognitive load for more complex PSA cases or other case scenarios. Furthermore, the selection of cases may have contributed to a significant improvement between the laceration case and the abscess case. Sedation for laceration repair may be more cognitively intuitive than for incision and drainage. The difference may be also experiential if the learners experienced the laceration scenario second.

Another limitation of our study is that residents may not focus on the same aspects of performance as faculty raters. We did not define faculty and resident self-assessment performance elements and therefore these may be looking at different aspects of clinical care. This, however, was not a primary goal of our study and is a potential area of future research.

Since our study population was not large enough to detect subgroup differences, it is possible that the benefit is greater for more novice learners. Subsequent research should define the optimal contexts and adjuncts to realize the maximal benefits of video-assisted self-debriefing.

CONCLUSIONS

In conclusion, emergency medicine residents receiving feedback on their performance via video-assisted self-debriefing improved their performance in simulated emergency department procedural sedation to the same degree as with standard faculty debriefing. Video-assisted self-debriefing is a promising avenue for leveraging the benefits of simulation-based training with reduced resource requirements. Future work should delineate the relevant content areas and supportive structures need to optimize its impact on learner competency.

We would like to acknowledge the staff at JUMP simulation that made this study possible including Toufic Khairallah, research coordinators Kim Cooley and Kelsey Balcer, our simulation specialists (especially Dustin Holzwarth), and our standardized participants.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10372/full

Data Supplement S1. Supplemental material.
Defining “Swarming” as a New Model to Optimize Efficiency and Education in an Academic Emergency Department

Jessica L. Perniciaro, MD\textsuperscript{1,2}, Anita R. Schmidt, MPH\textsuperscript{1}, Phung K. Pham, MS, MA\textsuperscript{1}, and Deborah R. Liu, MD\textsuperscript{1,2}

ABSTRACT

Background: Academic emergency medicine is a constant balance between efficiency and education. We developed a new model called swarming, where the bedside nurse, resident, and attending/fellow simultaneously evaluate the patient, including initial vital signs, bedside triage, focused history and physical examination, and discussion of the treatment plan, thus creating a shared mental model.

Objectives: To combine perceptions from trainee physicians, supervising physicians, nurses, and families with in vivo measurements of emergency department swarms to better conceptualize the swarming model.

Methods: This mixed methods study was conducted using a convergent design. Qualitative data from focus groups with nurses, residents, and attendings/fellows were analyzed using directed content analysis. Swarming encounters were observed in real time; durations of key aspects and family satisfaction scores were analyzed using descriptive statistics. The qualitative and quantitative findings were integrated a posteriori.

Results: From the focus group data, 54 unique codes were identified, which were grouped together into five larger themes. From 39 swarms, mean (±SD) time (minutes) spent in patient rooms: nurses = 6.8 (±3.0), residents = 10.4 (±4.1), and attendings/fellows = 9.4 (±4.3). Electronic documentation was included in 67% of swarms, and 39% included orders initiated at the bedside. Mean (±SD) family satisfaction was 4.8 (±0.7; Likert scale 1–5).

Conclusions: Swarming is currently implemented with significant variability but results in high provider and family satisfaction. There is also consensus among physicians that swarming improves trainee education in the emergency setting. The benefits and barriers to swarming are underscored by the unpredictable nature of the ED and the observed variability in implementation. Our findings provide a critical foundation for our efforts to refine, standardize, and appraise our swarming model.

Background

Emergency medicine in the academic setting involves a constant balance between efficiency and education. In the age of ever-increasing volumes and emergency department (ED) overcrowding, the tension between these two priorities becomes particularly palpable.\textsuperscript{1} The serial model of patient evaluation, in which the trainee physician evaluates patients first...
before presenting to a supervising physician, further compounds this issue. While this traditional model of patient care has been held as a criterion standard for trainee education and autonomy, it is inherently inefficient and does not allow for active bedside learning and conscious role modeling, concepts that have emerged as important for medical education in the emergency setting. Existing emergency medicine literature describes a number of methods to improve front-end operations in the ED, including immediate bedding of patients and placing physicians in triage. Although there is evidence to suggest that these processes improve efficiency, the issue of resident education in the emergency setting is seldom addressed within these interventions. One example of simultaneously integrating education and efficiency can be found in the family-centered rounds literature, which has demonstrated improved staff understanding of the medical plan without significantly increasing rounding time. This model allows for team reflexivity, which entails group reflection and adaptation before, during, and after a patient encounter. Compared to the inpatient setting, however, the ED treats the undifferentiated patient with whom there is no prior relationship or existing plan, which limits the comparisons that can be made to this well-established model of care.

With the competing goals of efficiency and education in mind, our pediatric academic ED underwent a sweeping process redesign in November 2016. During this intensive 5-day period, we created a new model which included quick sorting of patients into tracks, rapid bedding without traditional triage, zoning of physicians, and increasing bed turnovers by moving patients into treatment-in-progress and discharge areas. A key component of the redesign that addressed both front-end efficiency and trainee education was the concept of “swarming.”

Importance
Swarming is a novel model that facilitates the immediate evaluation of a patient simultaneously by the bedside nurse, resident, and senior ED physician (attending or fellow). The bedside nurse performs an intake assessment including initial vital signs, while the physicians begin a focused history and physical. At the end of this encounter a plan is discussed with both family and nursing, and electronic documentation and orders are initiated. This approach, which embraces team reflexivity and mirrors the one often used in the management of critically ill and trauma patients, helps to develop a shared mental model for optimizing communication, care, and learning.

Following the process redesign which included the implementation of swarming, we saw significant decreases in length of stay (LOS), door-to-provider times, and left-without-being-seen rates. However, due to the novelty of this approach and the lack of well-defined best practices, significant variability in receptivity among staff and implementation continued to occur. To eventually develop best practices and evaluate the effects of swarming, the first step was to characterize the variation that currently exists and understand any potential barriers to consistent implementation.

Goals of This Investigation
The objective of this study was to combine perceptions from physicians, nurses, and families with in vivo measurements of ED swarms to better conceptualize the swarming model. In doing so, we sought to answer the following questions: What factors do physicians and nurses perceive as barriers to swarming implementation? What is the degree of family satisfaction with swarming encounters? What are the interactions that occur during swarming and what are their durations? Given the novelty of this approach, having the ability to triangulate the quantitative characteristics of the swarm with the perceptions of clinicians and families was essential for a comprehensive understanding of the model.

METHODS
Study Design and Setting
We conducted a mixed methods study using the convergent design. This approach involves separate qualitative and quantitative data collection, wherein the two data sets are collected around the same time and one does not influence the other; the data sets are then independently analyzed and results are coalesced to attain a convergent interpretation of the phenomenon of interest. All aspects of the study were carried out at the Children’s Hospital Los Angeles ED, a tertiary care academic pediatric ED with over 95,000 visits per year.

Selection of Participants
Focus groups were conducted to collect qualitative data; this method was chosen with the expectation that interaction among the members of each group would enrich the information collected. For the focus group portion of the study, a purposive sampling strategy was used; individuals with experience relevant to the
swarming context were approached for participation to generate a content-rich sample. This included resident physicians, attending physicians, fellows, and nurses. Resident physicians were contacted by e-mail and asked to volunteer for focus group participation. Residents who were working in the ED at the time of the focus groups were also invited to attend. For the attending and fellow physicians as well as the nurses, focus groups were conducted during regular staff meeting times. All participants were consented prior to the start of each session and participation was completely voluntary.

There is currently no consensus on sample size conventions for mixed methods studies; it has been suggested, therefore, that the sampling strategy and size estimates should be based on the specific research question and should support feasibility. As a result, a minimum number of 30 swarms were deemed as a feasible convenience sample size for direct observation. Patients were enrolled during times when research staff were available, from mornings to early evenings. Patients that were excluded from this phase of the study included critical (Emergency Severity Index [ESI] level 1) patients, where it was not feasible to consent families prior to initiating evaluation and treatment. Non–English-speaking patients and families were also excluded, as the presence or use of an interpreter could change the durations of various aspects of the swarm.

**Measurements**

Separate focus groups (nurses, resident physicians, and attending/fellow physicians) were conducted to facilitate open and unhindered conversation about swarming. Focus groups were conducted in a semistructured format, with a research staff moderator there to guide discussion as needed. Because the principal investigator (JLP) is a physician, the decision was made to have non-clinical research staff serve as moderators to encourage neutrality and open discussion, therefore increasing the trustworthiness of the collected data. Each participant was given a handout with several prompts, including asking about general perceptions of swarming as well as benefits of and barriers to the swarming model (see Data Supplement S1, Appendix S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10388/full). Moderators probed participants as necessary to optimize discussion time. All focus groups were audio recorded and then transcribed, removing identifying information during the transcription process.

During the observational portion of the study, a research team member recorded the duration of key aspects of the swarming encounter using a multitimer digital stopwatch (see Table 1 for a list of these timed variables and associated definitions). In addition, it was noted whether electronic documentation was started and whether orders were placed at the bedside during the swarming encounter. Thereafter, research staff administered a brief questionnaire with the family, where they explained the new swarming model and then asked a few questions regarding perceptions of the encounter, including a 5-point Likert scale to measure overall satisfaction (see Data Supplement S1, Appendix S2). Research staff also collected general information about the visit from the patient’s electronic medical record, including ESI acuity level, LOS at which swarming was initiated, chief complaint, and final diagnosis. Our mixed methods study was approved by the Institutional Review Board at our hospital.

**Data Analysis**

To establish trustworthiness of qualitative findings, it was decided that the research team should have two coders: one

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Names and Definitions of Variables That Were Timed During Direct Observation of Swarming Encounter</strong></td>
</tr>
<tr>
<td>RN in room</td>
</tr>
<tr>
<td>Resident MD in room</td>
</tr>
<tr>
<td>Attending MD in room</td>
</tr>
<tr>
<td>Resident MD history</td>
</tr>
<tr>
<td>Attending MD history</td>
</tr>
<tr>
<td>RN vitals and assessment</td>
</tr>
<tr>
<td>Resident MD physical exam</td>
</tr>
<tr>
<td>Attending MD physical exam</td>
</tr>
<tr>
<td>Resident MD discusses plan</td>
</tr>
<tr>
<td>Attending MD discusses plan</td>
</tr>
</tbody>
</table>
clinician (JLP) with relevant clinical swarming experience, and one non-clinical research staff member (ARS) with qualitative coding expertise. Directed content analysis was used to analyze focus group data.19 This strategy begins with deductive coding, which integrates codes from relevant literature, followed by inductive coding, which involves the creation of de novo codes as the coder goes through the transcript line by line. For our purposes, an initial code list was derived from literature pertinent to swarming, including ED front-end operations, ED process improvement, and FCR.2,6,10,20,30

Deductive coding was conducted initially; short portions of the same focus group transcripts were independently coded using the literature-based code list, after which the coders met and built consensus. This process was continued in an iterative fashion until all transcripts were coded using the literature-based codes. Throughout the process the code list was edited to include definitions, and unused codes were ultimately deleted. Inductive coding was subsequently performed: all transcripts were again independently reviewed line by line and new codes were created when appropriate. The process of building consensus was repeated, with frequent meetings between the coders to review and reconcile all new codes. Any coding discrepancies were resolved through discussion with a third member of the study group, the senior author on this study (DRL). Once the code list was finalized, codes were grouped into themes and subthemes, and the frequency of each code within each transcript was recorded. The complete coding structure is available in Figure 1.

Descriptive statistics were used for all time duration data collected from direct observation of swarming encounters. Mean family satisfaction scores were computed.

RESULTS

Characteristics of Study Subjects
There was one attending and fellow focus group with 14 participants, which included nine attending physicians and five fellow physicians. There was one nurse focus group with seven nurses, including both long-term and traveling nurses. Two focus groups took place with resident physicians, with seven and four participants, respectively. All resident physicians were pediatric residents based at our home institution (external rotating residents were not able to be included due to scheduling). In all groups, there was a mix of staff that had experienced our previous serial model of care prior to the redesign and those who had not. The length of the focus groups varied from approximately 13 to 21 minutes.

For the direct observations, 39 swarming encounters were ultimately included in the analysis. The mean (±SD) age of patients in this group was 6.23 (±5.24) years (median = 5.02 years, range = 2 months to 17 years). Of all encounters, 21% were categorized as ESI level 2, 64% were ESI level 3, and 15% were ESI level 4.

Qualitative Results
The coding process yielded 54 unique codes, and the total coding structure (Figure 1) is summarized in the following five sections, with representative quotes from the focus groups included in Table 2.

Inherent Qualities of the ED. Participants in all focus groups discussed how attributes that are perceived as unique to the ED often influence their ability to swarm. Overcrowding and unpredictability, which are inherent to the ED, as well as the effort required to gather swarming participants, were noted as challenges. Furthermore, fatigue was stated as being a barrier to swarming, and both attending physicians and nurses mentioned that they are less likely to swarm at the end of their shifts. Additionally, all staff recognized technological considerations, such as the availability of computers for electronic documentation, and physician zoning, as they relate to the swarming process.

System Implementation. Given the ED context, participants noted the difficulty involved in creating and sustaining a new model like swarming. In light of the frequent rotation of new residents in the ED, all physicians recognized the need for setting clear expectations about individual roles prior to swarming. Staff unanimously expressed a desire for greater standardization, mentioning the need for further clarification on implementation. Participants from all groups mentioned that there are certain patient conditions and chief complaints that lend themselves best to the swarming model.

Variability. A common theme that emerged from all of the focus groups was the variability with which swarming is implemented. Participants noted variation in length of the swarming encounter, as well as...
Figure 1. Themes, subthemes, and codes as derived from qualitative directed content analysis. The frequency of each code is listed after code title, divided into attending/fellow (A), nursing (N), and resident (R) focus groups. Total number of participants of each type: 14 attending/fellow physicians, seven nurses, and 11 resident physicians.
variability in execution depending on the specific nurses or physicians who are involved. Specifically, resident physicians noted inconsistency in how each attending physician carries out the swarming model, with discussion of how this impacts their understanding of a swarming standard.

**Efficiency Versus Education.** Perspectives on the optimal balance of efficiency and education in ED swarms varied significantly from group to group. Although there was consensus among all focus groups that swarming reduced time to treatment and increased overall ED efficiency, all groups noted that nurses were often resistant to the swarming process. Nurses expressed that their efficiency was their ultimate priority and as a result noted that staying in the patient room for the entire swarming encounter was often difficult for them. Both nurses and attending physicians specified that nurses are hesitant to participate in swarming when it is led by a resident physician; these two groups agreed, furthermore, that when a full team swarm is not feasible, their preferred alternative is a nurse-fellow/attending swarm. For residents, on the other hand, the preferred alternative is a resident-nurse swarm.

Residents discussed two barriers to education, namely the potential limitations of teaching directly in front of the patient/family and the difficulty of discussing sensitive topics. However, residents also recognized educational advantages to swarming, including attending role modeling, bedside teaching, directed feedback, and learning how to approach patients from an emergency medicine perspective. Residents spoke about the importance of functioning autonomously and taking ownership of patient care but expressed dissonance concerning the extent to which swarming helped or hindered those priorities. All physicians, including both supervising and trainee physicians, acknowledged that swarming has helped to develop a consistent shared mental model of patient care through improvement of multidisciplinary team engagement and team communication.

**Outcomes.** There were a number of comments made in all focus groups about positive outcomes related to swarming. Staff noted that, in addition to the overall improvement in patient care, patients and families were more satisfied with swarming in contrast to the serial model of care. They also mentioned that swarming facilitates the provision of timely and

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Representative Quotations From Focus Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inherent qualities of the ED</td>
<td>“The busier the ED the less swarming, so early shifts or later in your overnight shift I find myself swarming but if it’s mid shift and it’s crazy, you can’t find everybody to swarm at the same time.” (attending/fellow)  “The swarming is great, and I believe it should continue, but I just don’t know if it’s possible to swarm on every patient, because it’s just not practical. It gets busy, and there’s more than one of us usually at a station … The attending can’t be in both places at the same time.” (resident)</td>
</tr>
<tr>
<td>System implementation</td>
<td>“There are a lot of residents, it’s hard for us to continue to orient people to the swarm, it’s also something that’s changing, and so there are a lot of residents who don’t understand that the beginning and point of the swarm is to get the initial, I call it the primary survey of the visit. It’s not to do everything.” (attending/fellow)  “And when we first started this process as well, everyone was very diligent … I think it’s just we’ve gotten, in the year, we’ve gotten away from both the logistics of it and then also the way in which we’re approaching the assessment.” (nurse)</td>
</tr>
<tr>
<td>Variability</td>
<td>“It depends on the time of day, depends on the physician, the nurse, depends on the patient.” (nurse)  “I also haven’t seen it consistently done. It depends on how busy it is, it also depends on the attending you work with. I think some do it more than others.” (resident)  “I like that variability. And I don’t know if it’s necessarily the intent, but the practicality makes it so, and I think it kind of addresses your autonomy vs. learning from mentorship and watching your attending.” (resident)</td>
</tr>
<tr>
<td>Efficiency vs. education</td>
<td>“I’ve been told that [residents] like how we talk to families, like even in a difficult patient setting or family interaction, even just what they need to do to follow up or reasons to come back. So, they have said they like that component of the swarm, seeing how we manage the patient differently than they would.” (attending/fellow)  “From a resident education perspective, we are being taught to actually think like emergency medicine physicians.” (resident)  “The focus of my education is still more on the recognition and … less so the efficiency and the throughput part of it, so maybe sometimes that’s the down side I see at my point of training, like having that opportunity to see the patient by myself and kind of think through it, rather than how fast we can get through it.” (resident)</td>
</tr>
<tr>
<td>Outcomes</td>
<td>“And we can get things started, we don’t have to now wait for orders, things are much more timely.” (nurse)  “Patients seem happier because they’re being seen faster, and they don’t have to repeat their story over and over again to multiple people.” (resident)  “I really appreciate the communication we have with the nursing staff in this model.” (resident)</td>
</tr>
</tbody>
</table>
accurate information to the family and that families appreciate not having to repeat information multiple times. Notably, there was consensus among the physician focus groups about the high degree of overall provider satisfaction with swarming.

### Quantitative Results

For those swarming encounters included in the direct observation portion of the study, the initial evaluation occurred at a mean (±SD) of 21.2 (±11.2) minutes after patient arrival (median = 18 minutes, range = 3-54 minutes). During the swarming encounter, on average, the nurses spent a mean (±SD) 6.8 (±3.0) minutes in the patient room, the attending/fellow physicians spent a mean (±SD) of 9.4 (±4.3) minutes, and the resident physicians spent a mean (±SD) of 10.4 (±4.1) minutes (for all the time durations of the key aspects of the swarming encounter, see Table 3 for the means and SDs and Figure 2 for the medians and interquartile ranges).

During the 39 observed swarming encounters, 67% included electronic documentation that was started by one of the physicians, 39% included orders that were placed at the bedside, 67% included direct communication from the resident to the attending/fellow, 67% included direct communication from the attending/fellow to the resident, 13% included direct communication from the resident to the nurse, and 31% included direct communication from the attending or fellow to the nurse. On the questionnaire completed with the family after the encounter, mean (±SD) overall satisfaction was 4.8 (±0.7) on a 5-point Likert scale.

### Synthesis of Results

The perspectives shared during the focus groups can be used to better understand and interpret the quantitative results and vice versa. A theme frequently mentioned in the focus groups was the variability with which swarming is implemented. This was supported by the quantitative data, particularly in the wide SDs of the time durations of key swarming aspects. Of note, in the qualitative data, the resident physicians specifically discussed how the attending or fellow physicians varied in their implementation of swarming. Such variations included whether the supervising physicians remained in the room throughout the complete evaluation, whether they started electronic documentation or orders, or whether the senior physician took the lead in determining or explaining the plan to the family or let the resident guide this final aspect of the swarm. In particular, these inconsistencies of the supervising physicians as perceived by the trainees may have important implications on the educational effects of swarming. Staff unanimously noted the need for greater standardization of the swarming model.

During the focus groups, it was noted that nurses sometimes hesitate to swarm and the nurses themselves expressed that their own efficiency is of primary concern. These results converge with the finding that nurses spent the least time on average in the patient rooms during swarming. Although all swarming team members started out in the room together, nurses were more likely to leave before the other clinicians.

The quantitative results showed inconsistent durations of time spent in discussion between members of the care team during swarms, which could be interpreted as markers of poor team communication. However, in the qualitative portion of the study, the idea emerged that swarming often results in improvements in team communication and the creation of a shared mental model of care. This may be explained by the simultaneous presence of all providers in the room who have increased knowledge of the assessment and care plan as a result of hearing each other speak directly with the family.

The availability of computers for electronic documentation was a factor discussed in the focus groups as being influential in the swarming model. This idea was reflected in the quantitative data, where the frequency of electronic orders placed at the bedside varied significantly from one swarm to the next.

Finally, the average level of family satisfaction with swarming encounters was notably high. This converged with the perspectives in the focus groups, where both nurses and physicians mentioned that families appreciate not having to repeat information

### Table 3

<table>
<thead>
<tr>
<th>Time Durations of Key Aspects of the Swarming Encounter (Minutes)</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN in room</td>
<td>6.8</td>
<td>3.0</td>
<td>2.0</td>
<td>13.7</td>
</tr>
<tr>
<td>Resident MD in room</td>
<td>10.4</td>
<td>4.1</td>
<td>2.9</td>
<td>18.9</td>
</tr>
<tr>
<td>Attending MD in room</td>
<td>9.4</td>
<td>4.3</td>
<td>2.9</td>
<td>19.2</td>
</tr>
<tr>
<td>Resident MD history</td>
<td>3.1</td>
<td>2.5</td>
<td>0.1</td>
<td>14.3</td>
</tr>
<tr>
<td>Attending MD history</td>
<td>1.7</td>
<td>1.7</td>
<td>0</td>
<td>9.5</td>
</tr>
<tr>
<td>RN vitals and assessment</td>
<td>3.1</td>
<td>1.8</td>
<td>0</td>
<td>9.1</td>
</tr>
<tr>
<td>Resident MD physical examination</td>
<td>2.1</td>
<td>1.6</td>
<td>0</td>
<td>6.3</td>
</tr>
<tr>
<td>Attending MD physical examination</td>
<td>1.2</td>
<td>0.9</td>
<td>0</td>
<td>3.8</td>
</tr>
<tr>
<td>Resident MD discusses plan</td>
<td>0.6</td>
<td>1.0</td>
<td>0</td>
<td>4.5</td>
</tr>
<tr>
<td>Attending MD discusses plan</td>
<td>1.3</td>
<td>1.5</td>
<td>0</td>
<td>7.7</td>
</tr>
</tbody>
</table>
multiple times and are generally satisfied with the swarming model.

**DISCUSSION**

We embarked upon this investigation to describe and further conceptualize swarming, a new model to improve both efficiency and education in the academic emergency setting. The unique challenge of balancing these priorities in the ED has given rise to multiple innovative teaching strategies, which have been described in previous literature. Correspondingly, the results of our qualitative data reflect a general consensus that swarming improves education. Trainee physicians specifically noted the benefits of bedside teaching, attending role modeling, and feedback. While the literature supports the advantages of these educational tools, there are also caveats. Bedside teaching should be active, role modeling must be done in a conscious manner where the trainee analyzes the behaviors they witness, and feedback should be learner-centered and ideally in real time. As a team-based model where teacher and learner concurrently provide patient care, swarming is poised to embrace these best practices.

While there are multiple educational advantages of swarming, these may come at the cost of decreased resident autonomy, while acknowledging that supervising and trainee physicians may have different perceptions of the amount of autonomy provided in a clinical setting. Indeed, as reported in one family-centered rounds (FCR) study, senior residents often perceive decreased autonomy, and specific strategies can be employed to promote the autonomy of the resident physician in this team model. Setting clear expectations and defining responsibilities prior to the patient encounter are strategies that emerged from our results and are consistent with published FCR findings.

Our results also showed a consensus amongst team members that swarming increased efficiency compared to the traditional serial model of care. A number of studies show increased efficiency with changes in front-end operations, mostly focusing on rapid triage and physician/provider in triage. At the same time, a systematic review showed no statistically significant decrease in LOS with team triage, and other studies show improvements only with certain patient acuity levels. Although this study on swarming was not designed to quantify an improvement in efficiency metrics, our findings showed an average time commitment of approximately 10 minutes for the physicians, suggesting that a substantial amount can

---

**Figure 2.** Median time duration for key aspects of the swarming encounter with interquartile range (IQR). Across the boxplots, circles are outliers (values greater than IQR × 1.5 + 75th percentile, or less than IQR × 1.5 – 25th percentile), and asterisks are extreme outliers (values greater than IQR × 3 + 75th percentile). “Attending MD” in this setting refers to either the attending or fellow physician who was functioning in a supervisory role during the swarming encounter.
be accomplished in regard to patient care, communication, and education in a relatively short time frame.

Several differences exist between the swarming model and front-end operational interventions. In many of the physician-in-charge or team triage models, the team that performs the initial assessment is typically not the definitive treatment team. \(^8,28,30,34\) When one provider sees a patient in triage, with a different provider for the remainder of the ED stay, this increases hand-offs and may also increase the diagnostic studies ordered by the initial triage provider. \(^29\) In the swarming model, the team responsible for the initial assessment is the primary care team for the patient.

Another difference is the extent to which trainee education is incorporated. The majority of front-end operations do not include trainee physicians, with the exception of one team triage study that showed overall improvement in quality and efficiency with both a junior and senior physician present. \(^27\) The swarming model was designed expressly with the intention of integrating trainee education.

Although involvement of trainee physicians in the ED has been associated with increased LOS, \(^2\) an observational study of FCR showed that teaching behaviors, outside of directly teaching the physical examination, did not have a statistically significant effect on rounding time. \(^10\) This may indicate that the serial model of care traditionally used in the academic ED setting—where many aspects of history, physical examination, and assessment are repeated at separate times by training and supervising providers—may be the cause of the increased LOS seen with trainee involvement, rather than the educational activities themselves.

One of the purposes of this investigation was to conceptualize swarming to establish a foundation and clarify future directions for research. A potential area of investigation involves best practices in swarming. This study will assist us in standardizing this model—which is critical for consistent training and education of staff. At the same time, there may be certain types of patients or complaints or certain environmental factors within the ED that lend themselves more toward successful use of the swarming model. With further research, we hope to delineate the circumstances where a more uniform swarming model would be appropriate and the conditions under which intentional exceptions or variations may be made.

We also plan to evaluate the effects of swarming in comparison to the traditional serial model of care with regard to efficiency and education. Since there are times in the ED when we are simultaneously utilizing swarming with some patients and traditional serial care with others, there exists an opportunity to directly compare these two models while other environmental factors are consistent. We plan to measure specific efficiency metrics such as door-to-physician time, time to orders, time to treatment, and overall LOS.

Finally, we plan to evaluate the effect of swarming on tangible educational outcomes. This study has highlighted specific areas of trainee education that may be particularly affected by swarming, such as role modeling, bedside teaching, and resident autonomy. In future investigations, we aim to measure these outcomes, gathering information to formalize educational best practices in the swarming model.

**Limitations**

For the observational portion of the study, swarms were conveniently sampled based on when a provider decided to swarm for a patient evaluation and when research staff was available; these encounters tended to occur during less busy times in the ED, most often in the morning. As a result, there was some selection bias in favor of visits that occurred during morning hours.

The decision was made to exclude non–English-speaking patients from the study because the presence of an interpreter could change the durations of various components of the swarm. This choice, however, resulted in the exclusion of Spanish-speaking patients/families, which compose approximately one-third of our ED patient population. The patient/family satisfaction questionnaire was completed with a research staff member present, which could have induced social desirability bias in families and therefore affected their responses.

For the qualitative portion of the study, focus groups were not of homogenous size due to scheduling constraints for certain participants. Additionally, the choice was made to have research staff moderate the nurse and resident focus groups, since moderation by the principal investigator (a physician) could have influenced responses. Having different individuals moderate the various focus groups may have led to slight differences in response types, although there were also benefits to having research staff moderate these groups, including more open discussion and subsequently more trustworthy qualitative data. Furthermore, it is possible that participants with strongly positive or negative opinions about swarming may have been more likely to volunteer to participate in focus groups, thereby introducing some bias.
Of note, due to scheduling constraints we were unable to include rotating resident physicians from other institutions so only pediatric residents based at our institution were involved. Residents from external programs, which include pediatrics, family medicine, and emergency medicine, may have brought different perspectives to the focus group about swarming, especially with regard to its educational impact. Despite best efforts to facilitate open conversation by all participants, we cannot exclude the possibility that members of the focus groups may have censored comments due to existing relationships with other participants.

CONCLUSIONS

Swarming was created as a novel model in an attempt to balance our clinical and educational missions, two priorities that are truly at the core of academic emergency medicine. This investigation sought to better conceptualize the approach by combining perceptions from physicians, nurses, and families with in vivo measurements of swarming encounters. We ultimately found multiple benefits and barriers, which are underscored by both the inherent variability in the ED and the observed variability in the implementation of swarming. The findings presented here provide a critical foundation for our future efforts to refine, standardize, and appraise our swarming model.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10388/full

Appendix S1. Questionnaires used in staff focus groups.

Appendix S2. Questionnaire used in brief family interview.
Showing Your Thinking: Using Mind Maps to Understand the Gaps Between Experienced Emergency Physicians and Their Students

Kira Gossack-Keenan, MD1, Kerstin De Wit, MBChB, MD, MSc, MRCP, FRCEM, FRCPC2, Emily Gardiner, MD3, Michelle Turcotte, MD Candidate4, and Teresa M. Chan, MD, FRCPC, MHPE, DRCPCSC2,5

ABSTRACT

Background: Clinical teaching faculty rely on schemas for diagnosis. When they attempt to teach medical students, there may be a gap in the interpretation because the students do not have the same schemas. The aim of this analysis was to explore expert thinking processes through mind maps, to help determine the gaps between an expert’s mind map of their diagnostic thinking and how students interpret this teaching artifact.

Methods: A novel mind-mapping approach was used to examine how emergency physicians (EPs) explain their clinical reasoning schemas. Nine EPs were shown two different videos of a student interviewing a patient with possible venous thromboembolism. EPs were then asked to explain their diagnostic approach using a mind map, as if they were thinking to a student. Later, another medical student interviewed the EPs to clarify the mind map and revise as needed. A coding framework was generated to determine the discrepancy between the EP-generated mind map and the novice’s interpretation.

Results: Every mind map (18 mind maps from nine individuals) contained some discrepancy between the expert’s mind and novice’s interpretation. From the qualitative analysis of the changes between the originally created mind map and the later revision, the authors developed a conceptual framework describing types of amendments that students might expect teachers to make in their mind maps: 1) substantive amendments, such as incomplete mapping; and 2) clarifications, such as the need to explain background for a mind map element.

Conclusion: Emergency physician teachers tend to make jumps in reasoning, most commonly including incomplete mapping and maps requiring clarifications. Educating EPs on these processes will allow modification of their teaching modalities to better suit learners.

Dual-process theory postulates that decision making relies on two types of thinking: heuristic and analytical, otherwise known as fast and slow thinking.1,2 Fast thinking (a.k.a. system 1) is automatic and intuitive and requires little effort (e.g., knowing that an animal is either a zebra or horse); slow thinking (a.k.a.
system 2) is effortful and process-driven and requires focus (e.g., explaining to a student how to arrive at an answer). While it has been shown that both trainees and faculty similarly engage in both fast and slow thinking, experience can lead novices to perceive jumps in diagnostic reasoning when listening to the explanations of those more seasoned.

Within the medical field, fast thinking can allow physicians to quickly make diagnoses. Conversely, medical students are relatively inexperienced clinicians; as such, they are more likely to engage in “slow thinking” because their heuristic patterns have not yet fully developed. These novices must learn both implicitly by observing experts making clinical judgments and by explicit teaching. In the clinical teaching environment, the two phenomena intersect. To effectively teach novices, at times the expert needs to deconstruct their reasoning process into analytic steps.

Expert physicians use system 1 thinking to make diagnoses; medical students observe and infer diagnostic thinking based on experts’ observable or stated reasoning. Unsurprisingly, students are often left confused when observing their expert teachers make diagnostic decisions. What appears obvious to an attending because of their previous experience and well-developed heuristics, seems like a leap of logic for the inexperienced apprentice. Teachers can use diagrams or drawings to explain their thinking to learners at the bedside.

Medical teachers use visual aids at the bedside to emphasize key teaching points to learners. Visual aids can take on many forms, ranging from short notes (“Post It Pearls,” which are often captured at the bedside on sticky notes) to full didactic white board minilectures. The use of visual aids (such as mind maps) allows portrayal of each discrete point which contributes to the diagnostic conclusion, thus showing the slow thinking learning points entailed in the physician’s final judgment. First developed in the 1970s, mind mapping is one such visual teaching resource. Mind mapping relies on the visual portrayal of information. Mind mapping can help learners to remember key points and quickly review information. Mind mapping also helps to make teaching points more accessible to learners, by portraying them in visually interesting ways, emphasizing only key words. Visual aids and mind maps may therefore provide an opportunity to augment traditional bedside teaching, by forcing teachers to clarify their clinical reasoning. While faculty use mind maps, handwritten algorithms, and diagrams to teach, it is not known if these are effective tools for novice student learners.

The objective of this study was to demonstrate the gap between faculty teaching aims and the student’s learning. To do so, we reviewed mind maps as a teaching aid; we asked our participants to generate mind maps initially based on our observed simulated teaching videos and then asked them to explain their own mind maps at a later date with the aid of a trainee research assistant who was seeking clarification of elements within their diagrams.

**METHODS**

A novel mind-mapping approach was used to examine how experienced emergency physicians (EPs) explain their diagnostic reasoning to learners, through a think-aloud interview process using video-prompted mind mapping. Figure 1 provides an overview of the entire study protocol. In this study, our EPs were asked to explain their diagnostic reasoning to a medical student via a mind map; the intent of our design was to examine the explicit cognitive processes they use to explain patient diagnosis during a simulated teaching setting. Example clinical cases suggestive of pulmonary embolism or deep vein thrombosis were created to elicit this reasoning, with presentations including chest pain, breathlessness, and leg symptoms. We chose to use this knowledge elicitation technique to go beyond the usual think-aloud interview protocol, to leave a teaching “artifact”, which would allow us to revisit their initial thinking process.

This study contained analysis of the mind maps that were created during these think-aloud interviews, as well as a planned, secondary analysis of the changes EPs made to their mind maps during a follow-up interview (member check reinterpretation). The aim of this analysis was to determine the gaps between an expert’s mind map of their diagnostic thinking (meant to portray the slow thinking underlying their reasoning) and a trainee’s ability to interpret these maps.

Both the mind maps generated in their first interview and the later amendments to mind maps underwent qualitative analysis by two investigators, KG and TC. The investigators reviewed the original interpretation and the member check reinterpretation iteratively, generating a coding framework for the types of changes that were made to the mind map interpretations. Coding disagreements were resolved through a consensus building procedure, and only three codes
required this process. Each of the spreadsheets was then recoded by a single investigator (KG) to generate a tally of the changes that occurred via the member checking process.

Participants
Study participants were university-affiliated staff emergency medicine physicians, across three academic hospitals in Hamilton, Ontario, Canada. We recruited a total of nine physicians from these hospitals, through a convenience sample. Participants had varying levels of clinical experience. Participant demographics are found in Table 1.

Procedures
A study protocol was written prior to the start of the study, and two research assistants (EG and MT) were trained to administer the protocol; EG and MT were both students at the time of administering the study protocol interviews (EG was a second-year medical student, and MT was a premedical student). Ethical approval for this study was received from the Hamilton Integrated Research Ethics Board (HIREB #15-246).

To ensure that they were familiar with the concept of mind mapping, EPs were first shown an example video demonstrating the construction of a mind map.
A nonclinical example was used, involving a car that will not start (e.g., car will not start → check if proper key → check for visible damage …). This scenario was chosen as a practical, real-life example to which EPs could relate, to demonstrate how mind mapping should present an approach to a problem. After watching this demonstrative example of mind mapping, EPs were shown a clinical video of a medical student taking a patient history. To capture a breadth of cases, a total of six patient interview videos were created (see Data Supplement S1, available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1111/ace.m.10379/full, for instructions for the participants and details about the cases). EPs were randomized to view two of these six possible videos. The scenarios in these videos were created with the consideration of venous thromboembolism (VTE) as a potential diagnosis, with varying degrees of clinical likelihood. The topic of VTE was selected since it is an area with many evidence-based guidelines, which might drive diagnostic procedures and would result in rich discussions.

After viewing each video, EPs were asked to explain their clinical approach to the scenario. They were instructed to explain how they would diagnose the cause of the patient’s symptoms by thinking aloud their decision-making process and drawing out a mind map.

Emergency physicians were explicitly advised to assume that the interviewer had no prior knowledge surrounding VTE diagnosis, to encourage them to fully explain their diagnostic reasoning using slow thinking. After watching the video, EPs were provided with pen and paper to create their mind maps. EPs were not explicitly instructed regarding which specific information to include in their mind map, to allow a true representation of their approach. They drew a mind map and explained approach to VTE to the research assistant. The explanation was audiotaped and transcribed for reference alongside the visual aids.

**Interpretation of Mind Maps.** An investigator (a senior medical student [KG], not involved in the initial interviews) independently reviewed the mind maps to break down each map into component steps (or individual teaching items). An example of this is shown in Figure 2. The method of interpretation and analysis was done in the style of Cristancho and colleagues,15 who used this method to analyze drawings by surgeons. To augment the interpretation of the mind maps, our investigatory team used the original session transcripts to better understand the mind maps that were originally generated. This investigator also had access to the transcripts of the discussions, to augment the visual artifacts generated by our participants. Table 1 shows two examples of mind map subcomponent analysis, chosen as illustrative examples. Each of the 18 mind maps was treated similarly.

**Member Check.** Our medical student investigator (KG) then re-interviewed each physician to confirm the inferred learning points items, to question meaning behind those items that were not clear and to provide an opportunity for EPs to amend, supplement or change their maps. Each interview was based on the mind map that the EP generated and hence consisted of specific targeted questions concerning elements of each mind map that were present. Specifically, EPs were given the research assistant’s interpretation and their original mind map and then asked if they agreed or not. If they confirmed agreement, the research assistant moved on, but if the query prompted a change, then the adjustment of the interpretation was written verbatim. This more structured member-checking process allowed us to determine which subcomponents were misinterpreted or required further clarifications or additions. During the re-interview stage our research assistant attempted to prompt EPs to engage in slow thinking teaching by asking key questions to clarify unclear mind map points. These questions were intended to make logical jumps explicit with slow thinking explanations, to augment learners’ clinical understanding.

### Table 1

<table>
<thead>
<tr>
<th>Demographics of Physician Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>7 (78)</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>2 (22)</td>
</tr>
<tr>
<td><strong>Type of emergency medicine training</strong></td>
</tr>
<tr>
<td>College of Family Physicians of Canada, Emergency Medicine Program</td>
</tr>
<tr>
<td>5 (56)</td>
</tr>
<tr>
<td>Royal College of Physicians and Surgeons of Canada Training Program</td>
</tr>
<tr>
<td>4 (44)</td>
</tr>
<tr>
<td><strong>Other descriptors</strong></td>
</tr>
<tr>
<td>Years in emergency medicine specialty training</td>
</tr>
<tr>
<td>3 (1-5)</td>
</tr>
<tr>
<td>Years as a practicing physician</td>
</tr>
<tr>
<td>18 (8-25)</td>
</tr>
</tbody>
</table>

Data are reported as n (%) or median (interquartile range).
RESULTS

We recruited nine physicians to participate in this phase of the study. Table 1 depicts the demographic data of our participants. Data were included from the nine EPs yielding a total of 18 mind maps (two from each EP). Although there was no set time limit, EPs generally spent less than 10 minutes creating each map. The member checking interviews lasted 30 to 40 minutes.

Qualitative Analysis of Amendment Types

By examining the coding tables generated by discussing and comparing in the member-check process, we were able to qualitatively classify the amendments made by the EPs when asked to clarify their mind maps. Tables 2 and 3 define the various types of changes, as pertaining to the redrawn diagram. The two major categories of changes were substantive deficits to the mind maps (e.g., incomplete, unclear, complete changes/revision) and changes that merely required verbal clarification (e.g., adaptations, clarifications around their writing).

Mind-mapping Results

Box 1 provides a detailed summary of the discussion types that occurred during the re-interview. As seen in the example within Box 1, participant 7 required the research assistant to prompt them for further clarification.

Observations From Mind Map Member Checks

Certain EPs were able to shift to a slow-thinking-based method of teaching upon review of their mind maps, while others remained unable to do so even with prompting. We inferred that the “fast-thinking”
<table>
<thead>
<tr>
<th>Main Category</th>
<th>Subcategory</th>
<th>Amendment Type</th>
<th>Definition of Amendment Type</th>
<th>Number of Type in Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantive deficits</td>
<td>Incomplete</td>
<td>Selective focus</td>
<td>Participants tended to expand more thoroughly in sections of the mind map on re-interview, e.g., would go more in depth on a particular part of the schema they previously drew, calling to attention a certain aspect.</td>
<td>4</td>
</tr>
<tr>
<td>Information missing from mind maps</td>
<td>Missing steps</td>
<td>Additional step added within the mind map, e.g., an additional diagnostic step such as ultrasound.</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing rationale</td>
<td>Additional info needed to explain rationale behind clinical steps (e.g., CXR done to rule out pneumothorax, trauma).</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing data</td>
<td>Adding additional data to map, e.g., adding additional factors to consider in patient’s likelihood of DVT.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing critical diagnosis</td>
<td>Adding critical diagnosis to differential diagnoses within mind map, e.g., pneumothorax in chest pain differential.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Missing diagnosis</td>
<td>Adding noncritical diagnosis to differential diagnoses within mind map, e.g., musculoskeletal pain in chest pain differential.</td>
<td>8</td>
</tr>
<tr>
<td>Unclear</td>
<td>Vague</td>
<td></td>
<td>Mind map required additional info to be added for clarity; e.g., original map stated ‘1st and 2nd line imaging,’ requiring explanation that this meant CXR versus CT.</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-universal short form</td>
<td>Using a nonuniversal short form that is unclear to the reader, e.g., “DVT” to denote the question of whether or not the DVT was provoked.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poorly constructed map</td>
<td>Map required modification because its construction was confusing to reader; e.g., one specific differential diagnosis was circled despite it being equally likely as all other differential diagnoses.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor placement of items</td>
<td>Changing the placement of existing items within the map, e.g., explaining that steps should be concurrent instead of sequential.</td>
<td>3</td>
</tr>
<tr>
<td>Complete change/revision</td>
<td>Removing step</td>
<td>Mind map step removed as either extraneous or incorrect.</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>Required verbal clarification</td>
<td>Adaptive techniques</td>
<td>Creating general rule</td>
<td>Applying a general rule to a clinical scenario; e.g., every patient with chest pain should receive an electrocardiogram.</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“What if?”</td>
<td>Extrapolating the current clinical scenario to an imaginary patient; e.g., if this patient had been pregnant, I would have done ‘x’ instead.</td>
<td>7</td>
</tr>
<tr>
<td>Clarifications required</td>
<td>Clarifications required</td>
<td>Required addition of background clarification, e.g., Other risks in history of DVT would include swelling, prior history.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clarifying/justifying reasoning</td>
<td>Adding a reason for clinical decisions, e.g., decision to anticoagulate would be based on a local thrombosis study.</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expanding priorities</td>
<td>Additional information added to clarify which information is most relevant, e.g., ranking list of differential diagnoses by likelihood.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expanding differential diagnosis</td>
<td>Adding additional differential diagnoses; e.g., PE and pneumothorax within respiratory causes of chest pain.</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expanding meaning of term</td>
<td>Adding additional information to established term; e.g., “examine leg” to mean looking for swelling, redness, bruising.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Logistic clarification</td>
<td>Explanation added for site-specific logistics; e.g., would anticoagulate the patient and have them return the next day as CT scans unavailable overnight.</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

CT = computed tomography; CXR = chest X-ray; DVT = deep vein thrombosis; PE = pulmonary embolism.
jumps manifest as gaps in the mind maps that are resistant to further prompting. An example of this phenomenon was participant 7, who refused to further clarify their mind map after multiple prompts stating: “All the info is in my map, I don’t need to add anything.” This may reflect an inability to switch out of their fast-thinking approach (i.e., they are skipping steps institutionally and unable to declare their thought processes), impacting the way in which they teach learners. Resulting mind maps thus displayed significant heterogeneity, both in learning points and in the extent to which EPs expanded these points.

**DISCUSSION**

In this exploratory mind-mapping study we found that experienced physicians tend to use fast thinking when diagnosing patients with symptoms of venous thromboembolism. We found that their teaching mind maps displayed fast-thinking–type links, rather than including the slow-thinking components we would expect to use when teaching inexperienced medical students. Experienced clinicians may generate incomplete
depictions of decision making, simply because they are unaware of the logical jumps (e.g., system 1–type heuristics) that they engage upon in their normal clinical reasoning.

Herein lies the tension between the duality of a clinician–teacher’s identity: whereas an expert clinician may develop and hone their diagnostic acumen, teachers need to be declarative and specific in their descriptions. Consider the work of Sibbald and colleagues, for instance, where they have shown that experienced clinicians can determine the clinical trajectory of a patient within a few minutes of observing the patient. Experts develop and hone their system 1 processes so that they can come to a decision about a patient’s illness, but when we ask these same individuals to concurrently teach junior trainees who need these processes unpacked and explained, our findings show that they stumble. Knowing that expertise in diagnostic processes and declarative knowledge (i.e. what an expert is able to say aloud) are not the same thing may be a crucial point within faculty development, since teachers may find it difficult to explain what it is that they see and think, and have other barriers to communicating their thoughts. These barriers may disproportionately affect inexperienced learners when using mind maps for teaching, suggesting that even the most junior of medical learners have an important role in prompting teachers to explain unclear “fast-thinking” teaching jumps.

Multiple studies have demonstrated mind maps to be a useful tool to improve knowledge retention among trainees. Mind maps are effective in helping students recall central ideas, integrate critical reasoning, and apply problem-solving skills. They are particularly effective when students have limited prior knowledge of the topic, as is the case with medical students. While the usefulness of mind maps as a study technique has been proven, less research has been done on their use as a teaching technique.

Mind mapping may prove especially useful within the fast pace of the emergency department, which can be challenging for learning as bedside teaching sessions tend to be focused and brief. However, mind map use may leave learners with unanswered questions, which should ideally be addressed promptly as learners are unlikely to have further opportunities to ask for clarification. Our findings suggest that learners should ask probing questions to augment their learning, as their overall understanding of mind maps (and underlying clinical reasoning) increased when given the opportunity to ask questions. However, it is also important for teachers to be proactive in illustrating their thinking to students, because in some cases learners may not have enough background knowledge to ask the right probing questions. The fast-paced learning environment, as well as the learner’s desire to be perceived as knowledgeable within a hierarchical system, can all act as barriers to the learner speaking up and asking clarifying questions.

During re-interviews, multiple mind map amendments were made to each drawing. Some changes were prompted, while EPs made others independently after recognizing that their maps were missing key points or unclear explanations. These changes highlight gaps between a teacher and trainee’s perspectives on the same material and offer an opportunity to educate teachers on their explicit teaching skills. Mind maps are helpful in making these unclear learning jumps explicit, because they illustrate each step of reasoning. For example, upon re-interview one EP noted that the majority of their mind map reasoning relied on Wells’ pulmonary embolism criteria, but they never mentioned this aloud or noted it in their mind map. Reexamination of their mind map afterward was helpful in identifying this gap.

This study demonstrates that in order to teach with visual aids such as mind maps, there must be an effective dialogue between teacher and student. Constructivist models of teaching and learning show that co-construction with trainees will allow them to link previously learned concepts to newer ones. As such, engaging with learners to identify learning gaps may reduce skipped steps and cognitive jumps. Furthermore, trainees must be empowered to prompt and ask questions, which help EP teachers to clarify and explain their thinking; it should be noted that all our participants were readily able to explain and adjust their thinking when asked.

Since learners engage in cognitive apprenticeship throughout their training, it is especially important for teachers to invite trainees into a discussion around visual artifacts (like mind maps) to ensure that the intended teaching points are communicated. Further faculty development research focused on training teachers to articulate their thinking or to develop new strategies for explaining their thinking may be prudent.
LIMITATIONS

Limitations of this study include the small number of participants, which makes it difficult to draw generalizations about all teachers. Of course, this study was performed as part of an ex vivo experimental design—and so flaws in the EP mind maps may have been purely an artifact of this design. Furthermore, the mind maps were reinterpreted a period of time after their initial interview, and as such, the memory of their exact thinking may have been incomplete and could have yielded some of the differences between the first-round interview and the later member check. This study included physicians at three hospitals in one Canadian city and therefore may not be generalizable to other centers with different teaching experiences and teacher training. A convenience sample of EPs was used; therefore, EPs level of experience may influence the degree to which each EP engages in “fast” versus “slow” thinking. EPs were not shown the clinical scenario videos again during the re-interview stage, which may have influenced the changes they made to their teaching points. Finally, although the technique of mind mapping is frequently used by our participants, none had formal training in this technique; some may have been more talented at explaining difficult concepts and generating robust mind maps and some may have found the process foreign causing the resultant errors or misperceptions during the member check.

CONCLUSIONS

When clinical teachers express their diagnostic thinking via mind maps, they may not fully represent all of their cognitive processes effectively within their diagrams. The use of mind maps provides a visual representation of expert thinking, but may be wrought with errors if the faculty are untrained in this technique. Meanwhile, trainees must be aware that they can use questioning to help elucidate gaps in their diagrams and engage teachers in amending and/or clarifying mind maps. The gaps we identified between a trainee’s interpretation and teacher’s visual representation of their thinking may be a starting point for effective faculty and trainee development.

The primary authorship team thanks the students who contributed to this research.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10379/full

Data Supplement S1. Instructions & Six Clinical Scenarios from Video Prompts.
It’s Just Math—Unless It’s Toxic!

Kayla Myers, PharmD1, Elisabeth Giblin, MD2, and Michele Zell-Kanter, PharmD3

ABSTRACT

Objective: The objective was to determine emergency medicine residents’ ability to perform pharmaceutical calculations.

Methods: A six-question needs-assessment survey of emergency medicine residents was conducted at a citywide conference in 2016. Residents performed simple pharmaceutical calculations and were queried regarding resources they typically use to aid in performing calculations.

Results: Fifty-three emergency medicine residents out of 110 attendees (48%) completed the survey (postgraduate year [PGY]-1 n = 27, PGY-2 n = 13, PGY-3 n = 8, PGY-4 or above n = 5). Nearly 80% (n = 42) of all residents responded correctly to at least four of six questions (PGY-1 = 70%, PGY-2 = 86%, PGY-3 = 88%, and PGY-4 and above = 100%). Sixty-five percent of PGY-1s, 50% of PGY-2s, and 75% of PGY-3s thought that it was very important to correctly perform basic calculations. Google and Up To Date were frequently used to assist with calculations. More than 70% of all residents utilize clinical pharmacy (PharmD) staff for calculating doses. PharmD services were available at all times in the majority of survey responses.

Conclusions: Emergency medicine residents performed poorly when completing pharmaceutical calculations. This may contribute to increased patient morbidity and mortality. Internet resource use increased as the years of training increased. PharmD services were extensively utilized by all resident regardless of years of training.

More than 250,000 deaths are reported in the United States every year because of medical errors making it the third leading cause of death after heart disease and cancer.1 Medication errors contribute to higher mortality and cost of health care.2 Lack of knowledge and proficiency in pharmaceutical calculations may contribute to medication errors.3

The difficulties practitioners encounter when performing calculations is compounded by inconsistencies in the labeling of drug strengths. Concentrations of drugs in solution are expressed most commonly as mass per unit volume (e.g., mg/mL), ratios (e.g., 1:1,000), and percentages. The ratio system is based on thousands and percentages are based on hundreds adding to the potential for confusion and increasing the numbers of medication errors.4

METHODS

All emergency medicine residents representing six training programs who attended a regional conference in September 2016 were invited to participate in a needs-assessment survey. This survey was approved by our institutional review board. A convenience sample of emergency medicine residents were asked six multiple-choice questions (Figure 1). Questions on the survey involved calculating the concentration of drug...
1. Hurricane® spray contains 20% benzocaine. How many grams of benzocaine are there in 1 mL?
   a. 0.002 g
   b. 0.02 g
   c. 0.2 g (correct)
   d. 2.0 g

2. What percent ethanol is in 100 proof whiskey?
   a. 0.5%
   b. 5.0%
   c. 50% (correct)
   d. 100%

3. How much epinephrine is there in 1 mL of a 1:1000 solution?
   a. 0.01 mg
   b. 0.1 mg
   c. 1 mg (correct)
   d. 10 mg

4. How do you convert mcg/mL to mg/L?
   a. Divide by 1000
   b. Multiply by 1000
   c. Divide by 100
   d. They are equivalent (correct)

5. How many mcg are there in 1 mg?
   a. 10 mcg
   b. 100 mcg
   c. 1000 mcg (correct)
   d. 10 g

6. How many mL of 1% lidocaine can you safely infuse in a 20 kg child for laceration repair using a dose of 4 mg/kg?
   a. 2 mL
   b. 4 mL
   c. 6 mL
   d. 8 mL (correct)

7. How important do you think it is to be able to perform basic pharmaceutical calculations?
   a. Very
   b. Somewhat
   c. I can look up the answer
   d. Not important

8. What resource(s) do you use to find this information?
   a. Google
   b. Up-to-date
   c. Wikipedia
   d. Phone app
   e. Textbook
   f. Other: ______

9. Would you ask a clinical pharmacist (PharmD) to help you with these kind of calculations?
   a. Yes
   b. No
      PharmD is not available all times

10. I am a:
    a. PGY1
    b. PGY2
    c. PGY3
    d. PGY4

Figure 1. The six questions about pharmaceutical calculations (correct answers marked in bold).
solutions expressed as a percentage, a ratio, and mass concentration; calculating the correct volume; and mathematical unit conversions. No statistical analysis was performed due to small sample size.

RESULTS
Fifty-three emergency medicine residents completed the survey out of 110 attendees (48%). The numbers of residents per training year included: PGY-1 \( n = 27 \), PGY-2 \( n = 14 \), PGY-3 \( n = 8 \), PGY-4 and above \( n = 4 \). The mean score was 4 of 6 for all participants. Question 2 concerning the percentage of ethanol in 100-proof whiskey was answered correctly by 96% of participants \( (n = 49) \).

Only 51\% \( (n = 27) \) chose the correct amount of epinephrine in 1 mL of a 1:1,000 solution. More than half of the participants were able to correctly answer questions involving mathematical unit conversions (question 4 = 70\%, question 5 = 86\%). Nearly 80\% \( (n = 42) \) of all residents responded correctly to at least four of six questions \( (PGY-1 = 19/27, PGY-2 = 12/14, PGY-3 = 7/8, and PGY-4 and above = 4/4) \). However, only 47\% of all residents were able to answer five of six questions correctly. The most common incorrectly answered question (question 3) required residents to convert epinephrine concentrations.

More than half of the participants thought that it was very important to correctly perform basic calculations. Google (55\%) and Up-to-date (52\%) were the most commonly utilized resources followed by phone applications (45\%) and Wikipedia (15\%). More than 70\% of residents consult clinical pharmacists for assistance in calculating doses. Clinical pharmacy services were available at all times in the majority of survey responses (72\%).

DISCUSSION
Medication miscalculations and lack of understanding of drug concentrations may result in medication errors.\(^3\) Wheeler et al.\(^5\) conducted a Web-based multiple-choice questionnaire in the United Kingdom to determine whether physicians who had completed training were familiar with medications expressed as a ratio, a percentage, or mg/mL. They found that physicians perform poorly when calculating drug doses and have poor understanding of the various ways in which the drug concentrations in solutions are expressed.\(^5\)

Glover and Sussmane\(^6\) evaluated the calculation skills of pediatric residents. They found insufficient mathematical calculation competency in all 3 years of residency based on a 10-question test. Their overall average test score was 65\%. The years of clinical training did not improve test scores. None of the 21 residents who completed the test identified any of three infusion-related errors. Seven residents made 10-fold dosing errors and one resident made a 1,000-fold dosing error.

Honey et al.\(^7\) studied the frequency and type of resident prescribing errors in a pediatric outpatient setting. They compared the accuracy of residents in pediatrics \( (n = 16) \), family medicine \( (n = 9) \), internal medicine \( (n = 4) \), and medicine/pediatrics \( (n = 9) \) training programs. Thirty-eight residents wrote 2,941 prescriptions during the study. The prescribing error rate was 5.88\% for all residents. Error rates were greatest for family medicine residents (11\%) and lowest for pediatric residents (4\%). The most common error type was medication overdose although the authors did not specify if a miscalculation led to the overdose.

Our survey of emergency medicine residents had similar findings to Wheeler et al.\(^5\) Difficulties occurred when calculating the correct dose, especially when the drug concentration units were not uniformly expressed. In question 3 the epinephrine concentration was expressed as a ratio and the concentration needed to be calculated. Only 51\% of all emergency residents were able to calculate this correctly.

Calculation errors resulted when drug concentrations were expressed as a percentage (questions 1 and 6). Although more than half of the residents answered question 1 correctly, this was the most common incorrectly answered question for both the PGY-2 and PGY-3 residents.

Contrary to the pediatric resident findings of Glover and Sussmane, emergency medicine residency classes performed better as their years of training increased for the majority (four of six) of correctly answered questions, i.e., PGY-1 = 70\%, PGY-2 = 86\%, PGY-3 = 88\%, and PGY-4 or above = 100\% in our study.\(^6\) This result could be attributed to increased Internet use as the years of residency training increased. While the years of experience may improve residents’ ability to perform calculations, our results illustrating only 47\% of residents were able to correctly answer five of six questions demonstrates inconsistencies with experience (PGY-1 = 52\%, PGY-2 = 43\%, PGY-3 = 75\%, and PGY-4 or above = 50\%). The most commonly
missed question (question 3, epinephrine concentration conversion) across all years of training suggests that drug concentrations expressed as a ratio are confusing for all and that the level of experience does not improve residents’ ability to perform such calculations.

Overall, we found that emergency medicine residents could not correctly perform these calculations. This inability was greatest when drug concentrations were expressed as ratios and percentages. These deficiencies have the potential to cause dosing errors that may result in patient harm and poor outcomes.

Our results of emergency medicine residents were similar to findings of practicing physicians (mean experience length = 14.8 years) in a similar survey of dosage calculations. Necessary skills for performing dosage calculations should be a priority in medical school coursework. Emergency medicine residents would benefit from additional training in calculating drug doses. Medication doses calculated using computer-assisted dosing programs (e.g., heparin, conscious sedation medication) may be overridden by the prescriber and are not error proof. Residents should consult clinical pharmacy services to assist in calculating dosages, check doses, and prevent dosage errors.

LIMITATIONS

Study limitations include the small number of survey responses and the variable sample size between the residency classes. Residents who are intimidated by calculations may have chosen not to take part. The survey was conducted in a controlled environment and the results may be an underestimation of the magnitude of calculation errors made by emergency medicine residents in a chaotic emergency department milieu. This was a convenience sample of residents from six emergency medicine programs and the results may not be representative of emergency medicine residents from other training programs.

CONCLUSION

Emergency medicine residents performed poorly on a survey of pharmaceutical calculations. Uncertainty existed when calculating the correct dose using drug concentrations expressed as both a ratio and a percentage. Improved performance trends and utilization of web-based resources were seen with increased years of residency training.

Standardization of drug concentrations in solution as mg/mL may mitigate errors resulting from calculating drug doses. Practicing calculations in a simulated environment may also identify difficulties so that residents can be remediated.

References

1. Makary MA, Daniel M. Medical error—the third leading cause of death in the US. BMJ 2016;353:i2139.
Successful Implementation of a Resident Liaison to Medical Students in Emergency Medicine Rotations

Jessica Bod, MD, Alina Tsyrulnik, MD, Ryan Coughlin, MD, David Della-Giustina, MD, and Katja Goldflam, MD

ABSTRACT

Background: As the role of the resident-as-teacher grows, some residents are inspired to develop themselves early as leaders in education while in training. We describe the successful implementation of a resident liaison (RL) to medical students in emergency medicine (EM) as a way to develop resident leaders in medical education.

Methods: This position was implemented to develop interested residents as leaders in medical education and to provide medical students access to an EM physician who is closer to their training level and may be more approachable than the clerkship director. RLs are mentored by the clerkship director and are involved in curricular programming and education research.

Results: This innovation has strengthened our student EM rotations and has provided residents with a unique opportunity to explore a career in medical education. Residents have made tangible contributions to our educational programming in this role and have pursued careers in medical education. The program has been recognized as a “best practice” by students and the school of medicine.

Conclusions: The RL initiative has conferred significant benefits to residents and medical students. Implementation of a RL program may benefit EM rotations outside of our institution and perhaps outside of the EM specialty.

NEED FOR INNOVATION

As the concept of the resident-as-teacher grows and expands, some residents are motivated to pursue opportunities as leaders in education while in training. These residents are interested in learning about and being exposed to the entire scope of a career in educational leadership. This creates a need beyond the previously published resident-as-teacher curricula engaging residents in bedside, small-group, and lecture-based teaching. These motivated residents need programs that engage them in educational leadership opportunities, curricular design, and education research so that they can develop as clinician-educators.

BACKGROUND

The role of the resident-as-teacher has increasingly been recognized and emphasized in graduate medical education.1,2 Nationally, in emergency medicine (EM) and across many specialties, resident-as-teacher curricula have been implemented to train residents to teach.3,4 On our review of the literature, a formal program focusing on the development of an EM resident...
as a medical education leader or director has not been previously described.

At our institution, a subset of residents has demonstrated superior engagement in teaching activities and an interest in an education leadership role during residency. Some aspire to conduct medical education research or to prepare for careers as medical educators. Many resident-as-teacher curricula that primarily focus on bedside teaching and lecturing may not adequately prepare residents for careers as academic clinician-educators who have the additional academic responsibilities of developing curricula, administrating programs, and engaging in scholarship. Previous published work has demonstrated that there is a dearth of programming for residents who wish to pursue careers as clinician-educators. Programs that have been implemented in the past have required significant financial and infrastructural support and may not be easily reproduced at other sites.

In addition, rotating medical students at the clerkship and subinternship level were eager for a resident as a near-peer educator and a “safe” point of contact to voice concerns about the rotation. Near-peer educators have been shown to more easily establish a social rapport with learners that promotes a safe and comfortable educational environment, reduced anxiety, and free exchange of ideas. We developed and implemented the role of the resident liaison (RL) to medical students in our EM clerkship and subinternship in 2015 to serve the needs of the students and to train residents to become leaders in medical education.

**OBJECTIVES OF INNOVATION**

The objectives were to provide EM residents with mentorship, research, and educational leadership opportunities and to provide medical students with a resident leader for mentoring, feedback, and expression of concerns.

**DEVELOPMENT PROCESS**

There are very few published model curricula in education leadership for resident physicians. Our RL program was developed within the framework of the key roles of the clinician-educator as a clinical teacher, curriculum developer, administrator, and education scholar. The importance of mentorship and role modeling in supporting career satisfaction and achievement is well documented. In medical education, role models and mentors play a key role in the development of professional identity and career aspirations. The mentoring relationship of the clerkship director and RL is one of the cornerstones of the program. We set out to create a program that would be sustainable, adaptable, and reproducible at other institutions. A summary table of the program is included as a reference in Data Supplement S1 (available as supporting information in the online version of this paper, which is available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10398/full).

The RL is selected by the EM program director and clerkship director among interested senior residents. The resident selected has already demonstrated engagement in medical student education and has participated in the medical education “area-of-concentration” and resident-as-teacher program in the residency. These activities provide the resident with a background in education theory, curriculum design, and opportunities to conduct bedside and small-group teaching. The position is held for 1 academic year. To help ensure specific focus on the medical students, the role is not assigned to a chief resident.

Our program recognizes the importance of near-peer education and the value of a near-peer educational leader to understand and relay student concerns. Importantly, the RL is not tasked with any grading responsibilities and does not participate in resident selection. Students are informed of this policy during the orientation to their rotation. This allows the RL to function as a “safe” person for medical students to approach with any concerns about the clerkship or subinternship. Anonymized concerns can then be relayed to the clerkship director without student fear of judgment or negative impact on grading.

**IMPLEMENTATION PHASE**

The RL program was implemented in 2015 and is ongoing. Each year the residency program director and clerkship director select a qualified resident who has participated in the education area of concentration. The RL participates in all areas of programming for clerkship students and subinterns in EM, including direct teaching responsibilities such as orientation workshops, didactic sessions, and procedural workshops. Some of these activities, such as orientation, are more formal and allow the RL to interact with students in large groups. The RL, with the clerkship director present, is given the opportunity to address overarching
concerns such as course objectives, student roles, and responsibilities and professionalism expectations. Others, such as the procedural workshops and didactic sessions, take place in small groups. The RL is encouraged to design and pilot didactic sessions with the clerkship director. The clerkship director also observes small-group teaching sessions and gives feedback on technique. Training for many of these responsibilities is done via participation in the education area of concentration which focuses on adult education theory, feedback, and curriculum design. The RL also undergoes formal training in educational debriefing as part of the simulation rotation in residency.

Administrative responsibilities include leading the resident and medical student education committee, recruiting residents to teach workshops and lectures, and organizing introductions between rotating subinterns and potential mentors. The resident and medical student education committee is composed of EM residents of all years who are interested in medical education. Led by the RL, the group is responsible for planning activities for the EM interest group, scheduling residents to teach small-group sessions for clerkship students and distributing the student schedules to faculty. The RL attends quarterly feedback meetings with clerkship medical students and meets individually with subinterns to receive feedback on the rotation. The RL also attends the monthly EM medical student and residency education committee meeting, gives updates about issues in the clerkship and subinternship, and discusses any problems that have occurred or may arise.

The RL meets regularly with the clerkship director to identify career goals and to help develop curricular initiatives. The clerkship director also serves as the faculty mentor for education research projects that RL wishes to pursue as part of the experience.

OUTCOMES

Since implementation of the program in 2015, the RL position has become a coveted role in the residency. Of the four residents who have served in the role, one has become a clerkship director, one has taken an academic appointment focused on point-of-care ultrasound education, and one will complete a fellowship in education leadership. In exit interviews, the RLs expressed the position helped to solidify their commitment to a career in medical education and provided valuable insight into the administrative aspects of educational programming. They appreciated the leadership experience; the opportunity to work with faculty on educational initiatives; and the ability to develop academic projects, teaching skills, and a teaching portfolio.

Students have also valued the program. During curricular improvement feedback sessions, they cited the benefit of having a near-peer for frank feedback and concerns as well as a designated “go-to” person from whom to learn more about the specialty of EM and our residency program. Our medical school has recognized the RL program as a “best practice” among clerkships in their clerkship review committee and notes that this implementation signals a superior level of departmental investment in medical student education.

The RLs have made a significant impact on the curriculum of our clerkship and subinternship via their own initiatives. Some of these include the overhaul of the clerkship student schedule in response to student feedback, implementation of a lecture series for clerkship students, and creation of a procedure workshop for subinterns. They have completed research projects in medical education and have published these and/or presented them at national meetings. The flow of energy and fresh ideas has allowed our clerkship and subinternship rotations to evolve and thrive in a department with many competing priorities.

REFLECTIVE DISCUSSION

The RL program has unequivocally benefitted our educational programming for students and has produced some promising leaders in EM education. Implementation of an RL program may be easily accomplished in EM residencies outside of our institution. It is important to be intentional about the responsibilities and the academic goals of the resident so that the RL can plan the year to achieve those goals. We believe that the focus on mentorship is key in guiding RLs in their academic and curricular pursuits and in providing tangible benefit to the residents who are volunteering to take on extra administrative responsibilities. Clerkship directors seeking to implement an RL program should seek buy-in from residency program leadership, as well as from the school of medicine and should facilitate inclusion of the RL in all aspects of curricular planning. As the popularity of the program has grown at our institution, it has proven to be sustainable. Interested junior residents are working with the RL in educating medical students to be selected as the RL in the future. This enthusiasm has, in some
part, been sustained by the ongoing support of the residency leadership.

Challenges of the RL program include the balancing of education responsibilities with clinical work and the inherent challenges of doing education research. Chief residents in our program have protected time to engage in administrative responsibilities and it may be helpful to similarly protect time for RLs going forward. It may also be beneficial to seek opportunities for the RL to engage in education administration at higher levels in the medical school, for example, to be invited to attend a clerkship directors’ meeting. Despite these challenges, the RL initiative has conferred significant benefits to the RLs, the medical students, the education leadership, and the rotations in general. We are committed to continuing this program to train the next generation of clinician-educators.

References


Supporting Information

The following supporting information is available in the online version of this paper available at http://onlinelibrary.wiley.com/doi/10.1002/aet2.10398/full

Data Supplement S1. Resident liaison summary.
Five Tips for Building a Successful Didactic Talk

Despite recent interest in varied teaching modalities, the didactic lecture remains a staple of emergency medicine education. Preparing and delivering a memorable didactic talk is no easy task. Whether you are speaking in front of a handful of colleagues in a conference room or hundreds of strangers in a packed auditorium, the following key strategies will help you achieve lecturing success.

**PASSION + PURPOSE = ENGAGEMENT**

Passion is contagious. If the audience senses how much you care about your topic, they will feel more invested. Choose a topic that you crave to learn more about or share with others. Even if you are assigned a topic, find something personally interesting within that topic.

Find a story to frame your talk. In an editorial on storytelling in medicine, Ofri stated “Medical caregivers are always telling stories because stories provide meaning to much of their working lives ... the compulsion to tell a story is largely motivated by the profound emotions kindled by the clinical experience.” We share stories all the time—we just happen to call our stories “cases.” Stories bring a sense of humanity to a scientific talk and engage clinicians more than a collection of facts. A commonly encountered pitfall is that “most presentations make the supportive media (i.e. the slides) the focus of the presentation without thought about the story or the delivery.” This is not to say you must present a case—but you should put the key points into a larger organizing framework.

**KNOW YOUR AUDIENCE**

Understanding the demographics of your audience is essential to delivering a successful talk. Adult learners tend to ask “Why do I need to know this?” before committing their attention and will be more engaged if they see how the knowledge can be put to use. Miller suggests that we progress from fact gathering and interpretation to demonstrating and integrating our knowledge into practice as we become experts. A primary audience of new interns may need an overview of a topic and a few key details, focusing on what to do when they see this in the emergency department (ED). A more seasoned audience will appreciate an exploration of the latest research with a summary of what to consider the next time they encounter it in the ED. A typical weekly EM didactic setting may include all levels of competence, so it is prudent to ascertain the details of your expected audience ahead of time.

In addition to differing levels of expertise, audience members have individual learning styles: visual, auditory, read/write, and kinesthetic. Although each learner has a tendency to prefer one learning style, a majority of learners are multimodal, preferring a combination of learning styles dependent on the content or situation. Presenting information in multiple formats will engage more of your audience. Even kinesthetic learning can happen in a large group lecture by incorporating audience participation or passing around models.

**BONES AND FLESH**

Once you know your topic and your audience, choose three to five key points that you want your audience to walk away with. Fit these into your organizational framework to create the skeleton of your talk. Start fleshing out the talk by adding details such as examples, diagrams, current research, and/or a case. This brainstorming process should begin the moment you are assigned the talk—keep a template available and add ideas as they come (e.g., notepad, index cards,

The authors have no relevant financial information or potential conflicts to disclose.
notes app, word doc on laptop). Gathering ideas often helps solidify key points and you may realize you need to change your structure. Don’t feel you need to include everything. The details you ultimately include will depend on the key points you are trying to emphasize. If you find great resources that you don’t have time or space to include, consider creating a handout listing these extra resources.

As you deliberate what to include, you should also consider how you will present it. Find ways to engage your audience through active participation, rather than passive listening, with techniques like think–pair–share or clarifying questions. Instead of summarizing the latest research, compare recommendations from the literature with the practice patterns of your audience. A commonly used strategy is to have the audience work through a case. By asking “What would you do?” you increase engagement by allowing audience members to relate their own experience, another strategy that resonates with adult learners. Small-group discussions, polling software, or a show of hands engages more audience members than calling out a single member of the audience. Alternatively, asking an expert to comment may provide nuance that group discussion or polling cannot. The education literature and the Internet are rife with suggestions to increase audience engagement; a full review of potential techniques is beyond the scope of this article. Finally, it is essential to consider the overall length of your talk. Evidence suggests that learners’ attention levels begin to fade approximately 15 to 20 minutes into a standard lecture and that only three to five items can be retained at once.

**PICTURES SPEAK 1,000 WORDS, YOUR TEXT SHOULDN’T**

Images on slides should support what you are saying, not repeat it. Less is more: keep your slides simple, with as few words as possible. Audiences prefer talks with more image-based slides. Avoid overusing gimmicks such as animations or video clips. “Exercise restraint and always keep these three words in mind: simplicity, clarity, and brevity.” This technique will resonate with the audience and as result will be more impactful. The quality of the images you choose represents the quality of your entire talk, so choose high-resolution images. Shutterstock has a vast library of high-resolution images, for a fee. Free sites include Flickr, Pixabay, Pexels, and Freepik. Never apologize for a poor quality slide—don’t include it. Find or create a slide that does not need to be accompanied by an apology. When presenting a case be sure to include all relevant clinical images (e.g., x-rays, ECGs) and a clear, uncluttered presentation of key data like lab values and vital signs. Highlight relevant abnormalities or key findings.

**PRACTICE MAKES PERFECT**

Now that you have a rough draft of your talk, it is time to practice. The first time you actually hear yourself speak should not be during the presentation itself! It is critical to rehearse multiple times. Initial solitary rehearsals should still be out loud: when reviewing slides, we process information much faster in silence than when we actually have to speak the words. Speaking aloud as you rehearse gives you a much more realistic sense of how long your presentation will take. Rehearsing aloud also allows you to practice enunciating tricky words, altering your tone to stress certain points, fine-tuning the transitions between topics, and cementing the timing of slide changes. Practice your blocking: where will you stand? When will you walk to the other side of the stage? What gesture best emphasizes your point? As you rehearse you may realize you need to add or delete content or change your slides. Additionally, familiarity with the room arrangement can help you plan your delivery. Activities that work when the audience is seated at tables in a small room may not work in an auditorium. Finally, consider videotaping your rehearsal lectures. This will allow you to assess and fine-tune your pace, voice, and posture—subtle details that have a large impact on your overall presentation. Undoubtedly, it is beneficial to do this well in advance, with time to practice, rather than the morning of your talk.

As your presentation solidifies, the next step is to deliver your talk in front of a mock audience. This will allow you to gauge their captivation, try out some “test material,” and ensure you end on time. Choose colleagues who will provide you with brutal, open, honest feedback. Incorporate their suggestions and rehearse again. As Dr. Scott Weingart says, “The mark of an amazing presenter is that they’ve said the words countless times before … every presenter that seems so at ease, so capable of ad-libbing, of ‘never preparing their talk’ has probably prepared more extensively than you’d ever be able to imagine.”

Your mantra should be: Present … Reflect … Perfect
Repeat. If possible, arrive early to ensure that your presentation is loaded and displays properly. This is critical if your presentation is being loaded onto another computer. During your talk, remember to speak loudly, slowly, and clearly, which is easier to do if you’ve practiced.

If possible, have someone record your talk. It is also wise to obtain feedback from your audience. Review both and reflect on how you might further improve your presentation. Little nuances such as stage presence, posture, voice, tone, transitions, and cadence separate the outstanding from the average. Finally, after investing this much time and effort, don’t make your presentation a one-off. Look for opportunities to present again.

CONCLUSION

In the era of TED talks, a successful presentation is no longer about the slides but rather the unification of genuine passion, poignant stories, and stunning visuals that leave your audience moved and inspired. Top-notch presentations don’t just happen; they require planning, preparation, and practice. Follow these steps to give your learners the unforgettable presentation they deserve.

The authors have no relevant financial information or potential conflicts to disclose.

Eric Steinberg, DO, MEHP
Doug Franzen, MD, MEd

(franzend@uw.edu)

1Department of Emergency Medicine, St. Joseph’s University Medical Center, Paterson, NJ
2Department of Emergency Medicine, University of Washington School of Medicine, Seattle, WA

Supervising Editor: Wendy C. Coates, MD

References

In Reply to: Applicant Attitudes Toward the Association of American Medical Colleges’ Standardized Video Interview

In “Applicant Attitudes Toward the Association of American Medical Colleges’ Standardized Video Interview,” the authors describe their own program’s applicant experiences with the EMSVI. The viewpoints of the applicants toward the SVI were largely negative. While the authors do identify the limitations of their study, we as members of the AAMC Standardized Video Interview Work Group believe that additional context is important to consider. The data they published add to the body of literature around applicants’ experience with the SVI. In summer and fall 2017, the AAMC surveyed everyone who took the SVI. The first survey was administered immediately after applicants completed the SVI, and the second survey was administered after SVI scores were released. Both surveys addressed applicants’ experiences taking the SVI and their attitudes toward it. Results from the surveys had a much larger sample size, with each exceeding 2,000 respondents; users were found to have more mixed, but more positive, perceptions of the SVI. Consistent with the authors’ results, it found that applicants had generally negative reactions toward the SVI. They were skeptical of its ability to assess the target competencies and add value to the selection process. However, applicants also had more positive attitudes about the procedural aspects of the SVI (e.g., clarity of instructions, time allotted to respond) and preparation materials.

The authors’ main conclusion drawn in the paper is that applicants believe the SVI did not add value to their applications; the vast majority do not support its use. These findings are not particularly surprising, given that applicants already feel overwhelmed with the residency application process. They are also consistent with the employment literature, which shows that applicants have generally negative attitudes toward structured interviews and video-based interviews.

However, we must question what role these data should play in the national discussion of how we should be using novel selection tools in residency selection. Applicant reactions about issues of policy, procedure, usability, and preparation are very helpful because they can inform us of ways to improve the tool itself or educational needs. The best people to judge whether the tool is an accurate reflection of applicants’ competencies and whether it adds value to the selection process, however, are the program directors. Recently published evidence shows that the SVI does offer perceived value to program directors, which also offers a stronger argument for content validity.

What validity evidence does student satisfaction and perception bring to the discussion? Is student perception of the SVI more important to consider than their opinion on the role of other components of their application, such as USMLE scores or the personal statement? The authors could make a stronger argument as to why student subjective opinion of the SVI is uniquely valuable and why it should inform its inclusion in the selection process.

Regarding applicant concern about bias that the authors uncovered in this paper, data do show that there are no differences in SVI scores by race or ethnicity. However, it is valid to question how program directors viewing the videos may introduce bias themselves. The SVI, along with existing application...
components such as the candidate photo, letters of recommendation, and even clerkship grades are all potential sources of unconscious bias, and the medical education community should be examining these threats as well. AAMC provides training, including unconscious bias training, to program directors who signed up for the SVI pilot. In addition, it offers programs that wish to conduct a blind review of applications the ability to “turn off” SVI videos. This paper raises questions that should be asked about the value and limitations of all aspects of the residency application we are using right now.

Finally, the authors suggest that use of the SVI is an inevitability, although the AAMC has clearly indicated that the SVI is a pilot process. We will need to closely follow the results of the multiyear study further testing the validity of the SVI, currently under way as a partnership between the AAMC and 17 emergency medicine residency programs.

Applicants are certainly important stakeholders in the residency selection process, and their perspectives are valuable. However, the main conclusion we draw from the work of Winfield et al., is that we need to continue efforts toward demonstrating the validity evidence we do have for the SVI particularly by following matched residents over time in their residency training. Most importantly, we in the medical education community need to be applying the same scrutiny the SVI is undergoing to our historical application tools as well.

Nicole M. Deiorio, MD
Virginia Commonwealth University School of Medicine
Richmond, VA

References
Gender-based Harassment of Emergency Medicine Trainees: What Faculty Educators Need to Know

Dawn Jung, MD

ABSTRACT
There is increased interest in workplace harassment as reports of improper workplace conduct have surfaced among multiple industries and governments. Accounts of sexual misconduct in academic medical centers also highlight the need for further education and training for faculty educators. The purpose of this paper is to provide faculty educators with the tools to recognize and respond to gender-based harassment of medical trainees. More specifically, we will review existing literature, the definition of gender-based harassment, federal laws, institutional reporting protocols, and retaliation toward reporters of harassing behavior.

Accounts of sexual harassment in the workplace have surfaced in multiple sectors, including government and academia. Academic medical centers are not immune to the increasing number of reports of sexual harassment, highlighting the need for further education and training. The unpredictable environment of the emergency department can lead to an increased risk of trainees experiencing gender-based harassment. Trainees in the emergency department are particularly vulnerable due to the high volume of patients they may be meeting for the first time, who can possess varying pathologies and decision-making capacities. Gender-based harassment of trainees leads to emotional health problems, cynicism, thoughts of wanting to leave the specialty, and normalization of the experience. These negative effects of harassment on trainees can impair performance and ultimately affect patient care. Barriers that can prevent trainees from reporting harassing behaviors include shame, lack of confidence that they will be helped, and fear of negative consequences.

This paper attempts to provide faculty educators in the field of emergency medicine tools to recognize and intervene on instances of gender-based trainee harassment. The term gender-based harassment was chosen to include behaviors that are not overtly sexual, but still target a particular gender. Faculty educators are responsible for not only demonstrating workplace behavior free from harassment but also recognizing and responding to aberrant behavior toward trainees. Faculty can also have a role in helping trainees formulate response strategies to sexual harassment, which can be useful throughout a physician’s career.

EXISTING LITERATURE
In 1995, a landmark study by McNamara et al. drew into focus the high prevalence of gender-based harassment experienced by emergency medicine residents. For example, of 1,774 surveyed emergency medicine residents, 63% of women and 32% of men reported unwanted sexual advances, and 66% of women and 27% of men reported discomfort from sexual humor. This study further elucidated that the most common source of gender-based harassment for emergency medicine trainees were patients and family, followed by residents, nursing, and faculty. Since the publication of this study nearly 25 years ago, very few
additional studies have examined the incidence and source of gender-based harassment of emergency medicine trainees. Additionally, more recent studies lack the degree of detail necessary to discern what surveyors and responders consider to be sexual harassment. Survey questions that ask about sexual harassment in general, without discerning the frequency of specific types of behaviors, do not accurately estimate the prevalence of gender-based harassment. Even fewer publications address the steps faculty can take to help mitigate instances, or the negative effects of, sexual harassment of trainees. For example, in a recent landmark publication by the National Academies of Engineering, Science, and Medicine on sexual harassment of women, the closest recommendation specifically for faculty educators was general bystander training.

DEFINITIONS

Gender-based harassment, gender discrimination, and sexual violence have distinct and overlapping characteristics. For medical professionals working with trainees who may be targets of harassing behavior, it is important to recognize the distinguishing characteristic of each type of behavior. Gender-based harassment is any offensive and unwelcome conduct that targets a particular gender. Instances of gender-based harassment can include offensive comments that are sexual or gender-based, unwanted attention, offensive body language, and touching. Gender-based harassment is unlawful if severe and pervasive enough to create a hostile work environment or if the situation is considered quid pro quo harassment. A hostile work environment is created when harassing behavior is severe and pervasive enough to affect any reasonable person’s ability to do their job, limits opportunities for job advancement or receiving of benefits. Requesting any sexual or gender-based favors as a condition of employment or career advancement is a type of gender-based harassment, termed quid pro quo harassment. An example of quid pro quo harassment can be the following scenario: A resident invites a student out and the student declines. The resident facetiously responds that it might help the student’s grade. This scenario illustrates an instance of an offer of a favorable evaluation in exchange for the acceptance of an invitation.

In comparison, gender-based discrimination in medical education is any difference in the training environment between men and women. For example, discouraging male trainees from participating fully in obstetrics and gynecology cases creates a difference in the training environment for men and women and could be considered gender-based discrimination in medical education. In addition, the reluctance of patients to recognize female trainees as physicians could lead to a difference in training environment for male and female trainees and contribute to gender-based discrimination in the medical education setting. Sexual violence is any unwanted sexual experience and can be contact or noncontact. For example, being exposed to unwanted sexual attention, sexual comments, and exposure to explicit images can be examples of noncontact sexual violence. Of note, the previous examples are also instances of sexual harassment in addition to sexual violence.

FEDERAL AND INSTITUTIONAL POLICIES

For faculty educators to effectively support trainees, it is helpful to understand federal laws regarding gender-based harassment. Title IX of the Education Amendments of 1972 requires educational institutions to investigate and respond to any reports of gender discrimination, sexual harassment, and sexual violence. A recent ruling by the U.S. Court of Appeals in 2017 recognized residency programs as a type of educational program; thus, these programs are required to follow Title IX mandates. Since this recent ruling, it is unclear to whom the role of mandatory reporting of gender-based harassment will be assigned in the hospital setting. Federal laws recognize responsible employees as “any employee that has the authority to take action to redress sexual violence” or “whom (a trainee) could reasonably believe has this authority or duty.”

Gender-based harassment also violates Title VII of the Civil Rights Act of 1964. Title VII of the Civil Rights Act of 1964 prohibits employment discrimination based on race, color, religion, sex, and national origin. Hence, a residency program could be held accountable to both Title IX and Title VII during the handling of a case of trainee harassment.

INTERVENTION

Each academic medical center is unique in the means by which they respond to complaints of gender-based harassment. Two commonly recognized frameworks for institutions to address complaints of gender-based harassment is the grievance approach and the ombudsperson approach. The grievance approach is a
common reporting protocol that mimics the approach of most human resources departments. Through the grievance approach, the person(s) affected file a grievance with the supervisor of their respective education program or hospital department and the complaint goes through a chain of command. Another type of approach is the ombudsperson approach, where an appointed faculty at the institution is tasked with responding and resolving a variety of problems, including reports of gender-based harassment. Each approach has inherent weaknesses and strengths, the discussion of which is beyond the scope of this paper. When intervening or mentoring on behalf of trainees affected by harassment, faculty may consider if the reporting protocol at the institution more resembles a grievance approach or an ombudsperson approach. Additionally, becoming familiar with the protocol for resolution of a complaint and the methods by which involved parties will be counseled or reprimanded will help create effective counseling for trainees.

Faculty may check in and consult with the trainee that was the target of harassment, but respect the trainee’s confidentiality and autonomy as much as possible. Effective mentoring of a trainee includes assuring the faculty’s commitment to the trainee’s confidentiality and the limits of that confidentiality. Faculty educators, although controversial and institution dependent, may be considered mandatory reporters of gender-based harassment. Due to the potential sensitive nature of conversations regarding instances of workplace harassment, a phone call or in-person meeting is the best first mode of communication rather than detailed electronic exchanges. In addition, any electronic exchange between a trainee and faculty could be examined during the course of an investigation.

Of special mention are situations where patients and family are the source of trainee harassment. For physician educators, duty to treat patients can come into conflict with duty to educate and minimize threats to the learning environment. Faculty can navigate this difficult predicament by first assessing clinical acuity. After assuring the patient’s clinical stability, faculty educators can next take into account patient’s decision-making capacity. If a patient is medically stable and demonstrates decision-making capacity, consider counseling the patient regarding the effects of their behavior on the targeted trainee and healthcare environment. In addition, it is important for the physician educator to debrief and empower the trainee to determine the next steps. Finally, a decision can be made with the trainee on whether to accommodate, negotiate, or consider releasing a stable patient from care.

RETALIATION

Effective mentoring of trainees affected by harassment may include discussing potential retaliation scenarios. Retaliation for reporting workplace harassment is against the law. Examples of retaliatory behaviors can include ostracism, negative evaluations, negative recommendations, and hindering advancement. However, trainees can also be vulnerable to covert retaliation, which is difficult to prove and prevent. Binder et al. uses the term covert retaliation to describe vindictive comments made by a person accused of sexual harassment about his or her accuser in a confidential setting, such as grant review, award selection, or search committee.

Depending on the egregiousness and pervasiveness of harassing behavior, one option to discuss with the trainee is delaying making a report until the trainee is no longer under the perpetrator’s supervision and performance evaluations are complete (e.g., end of specialty or site rotation). Throughout the mentoring process, faculty can play an important role in ensuring that trainees maintain a sense of autonomy and control in decision making.

CONCLUSION

If faculty educators do not actively take steps to address harassing behavior, harassment will continue to persist. Trainees in emergency medicine are at increased risk of gender-based harassment due to the unpredictable and acute encounters with patients, cotrainees, staff, and faculty. Recognizing and effectively responding to harassing behavior takes practice, and faculty educators working with emergency department trainees should give themselves credit each time this is practiced.

References


17. Manning M. What if a resident or medical student is raped? Hospitals’ and academic medical centers’ Title IX obligations. AMA J Ethics 2018;20:3–9.


