

## EDUCATIONAL ADVANCE

# Critical Appraisal of Emergency Medicine Education Research: The Best Publications of 2014

Lalena M. Yarris, MD, MCR, Amy Miller Juve, EdD, Wendy C. Coates, MD, Jonathan Fisher, MD, MPH, Corey Heitz, MD, Philip Shayne, MD, and Susan E. Farrell, MD, EdM

## Abstract

**Objectives:** The objective was to critically appraise and highlight rigorous education research study articles published in 2014 whose outcomes advance the science of emergency medicine (EM) education.

**Methods:** A search of the English language literature in 2014 querying Education Resources Information Center (ERIC), PsychINFO, PubMed, and Scopus identified 243 EM-related articles using either quantitative (hypothesis-testing or observational investigations of educational interventions) or qualitative (exploring important phenomena in EM education) methods. Two reviewers independently screened all of the publications using previously established exclusion criteria. Six reviewers then independently scored the 25 selected publications using either a qualitative or a quantitative scoring system. Each scoring system consisted of nine criteria. Selected criteria were based on accepted educational review literature and chosen a priori. Both scoring systems use parallel scoring metrics and have been used previously within this annual review.

**Results:** Twenty-five medical education research papers (22 quantitative, three qualitative) met the criteria for inclusion and were reviewed. Five quantitative and two qualitative studies were ranked most highly by the reviewers as exemplary and are summarized in this article.

**Conclusions:** This annual critical appraisal series highlights seven excellent EM education research studies, meeting a priori criteria and published in 2014. Methodologic strengths in the 2014 papers are noted, and current trends in medical education research in EM are discussed.

ACADEMIC EMERGENCY MEDICINE 2015;22:000-000 © 2015 by the Society for Academic Emergency Medicine

Educators in emergency medicine (EM) have traditionally relied on experience and educational theory to guide teaching practice when faced with educational questions and problems. However, efforts to promote quality EM education research have increased over the past decade, highlighted by the 2012 *Academic Emergency Medicine* consensus conference, "Education Research in Emergency Medicine: Opportunities, Challenges, and Strategies for Success"; faculty

development efforts, such as the Medical Education Research Certificate (MERC) at Council of Emergency Medicine Residency Directors (CORD) program; and increasing grant opportunities from the Society of Academic Emergency Medicine (SAEM), CORD, and Emergency Medicine Foundation.<sup>1-4</sup>

In this seventh installment of the annual critical appraisal series, we used previously published criteria to critically analyze and rank the EM education research from

From the Department of Emergency Medicine (LMY) and the Department of Anesthesiology & Perioperative Medicine (AMJ), Oregon Health & Science University, Portland OR; the Department of Emergency Medicine and Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, University of California, Los Angeles David Geffen School of Medicine (WCC), Los Angeles, CA; the Departments of Emergency Medicine, Maricopa Medical Center & University of Arizona COM-Phoenix Campus (JF), Phoenix, AZ; the Department of Emergency Medicine, Virginia Tech Carilion School of Medicine (CH), Roanoke, VA; the Department of Emergency Medicine, Emory University School of Medicine (PS), Atlanta, GA; and Partners Healthcare International and Department of Emergency Medicine, Harvard Medical School (SEF), Boston, MA.

Received July 12, 2015; accepted July 22, 2015.

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

Drs. Yarris, Coates, and Heitz, who are associate editors at this journal, had no role in the peer review process or publication decision for this paper.

Supervising Editor: John Burton, MD.

Address for correspondence and reprints: Lalena M. Yarris, MD, MCR; e-mail: yarrisl@ohsu.edu.

2014.<sup>5</sup> The focus of this article is to review and highlight the methodologically superior studies that are pertinent to teaching and education in EM. Trends in EM education research over the past 7 years, as they can be inferred from this review, are summarized. It is hoped that this paper will serve as a valuable resource for EM educators and researchers invested in the scholarship of teaching.<sup>6</sup>

## METHODS

### Article Identification

We applied a previously described search strategy to identify all education research publications relevant to EM education.<sup>5</sup> Publications were limited to English language papers published in 2014. Searches were run in February 2015.

### Inclusion and Exclusion Criteria

Publications relevant to the EM education of medical students, residents, academic and nonacademic attending physicians, and other emergency providers were included. Medical education studies were defined as hypothesis-testing investigations, evaluations of educational interventions, or explorations of educational problems using either quantitative or qualitative methods. Publications were excluded if: 1) they were not considered to be peer-reviewed research (such as opinion pieces, commentaries, literature reviews, curricula descriptions without outcomes data), 2) they were not relevant to EM learners (such as reports on education of prehospital personnel and international studies that could not be generalized to EM training outside of the country in which they were performed), 3) single-site survey studies, and 4) studies that examined outcomes limited to an expected learning effect without a comparison group.

### Data Collection

Two authors independently screened 243 abstracts from all retrieved publications and applied the exclusion criteria. All differences in opinion were resolved by discussion. Retrieved publications were maintained in a Microsoft Excel 2010 database. Twenty-five publications were made electronically available for all six reviewers to score independently.

### Scoring

The publications were first assigned to a scoring system based on whether they were primarily quantitative or qualitative studies. The quantitative studies used scoring criteria, developed in 2009, and then continually optimized and iteratively modified since then.<sup>5,7-10</sup> Quantitative studies were scored in nine domains for a maximum total score of 25 points. The domains included the following: introduction (0-3 points), measurement (0-4 points), data collection (0-4 points), data analysis (0-3 points), discussion (0-3 points), limitations (0-2 points), innovation (0-2 points), generalizability (0-2 points), and clarity of writing (0-2 points). Each of the domains was scored based on predefined criteria to make scoring as objective as possible.

Using accepted recommendations and hierarchical formulations,<sup>10-12</sup> qualitative studies were assessed and

scored in nine domains, parallel to those applied to the quantitative studies, for a maximum total score of 25 points. These also included the domains of measurement, data collection, and data analysis criteria, as defined specifically for high-quality qualitative research. The scoring criteria for both quantitative and qualitative research studies have been previously published in this review series and are presented in Tables 1 and 2.<sup>5</sup>

### Data Analysis

Reviewers were excluded from scoring publications in which there was deemed to be significant conflict of interest (own publication, own institution, or vested interest in the authors or work). Publications were listed alphabetically by first author surname, and each reviewer started the review with the article whose author surname was alphabetically closest to that of the reviewer's surname. This process has been used throughout this review series in an attempt to mitigate bias resulting from reviewer fatigue. Each reviewer independently reviewed and scored each publication, except those excluded for conflict of interest. A total rating score was calculated for each article and entered into a spreadsheet using Microsoft Excel 2010. Using each reviewer's total rating score for each article, a rank list of quantitative studies and a rank list of qualitative studies was created for each reviewer. The rankings were then averaged among all six reviewers to prevent overvaluing any one reviewer's scoring. The a priori criteria for quantitative studies to be featured as exemplary were: 1) the average of all reviewers' rankings of an article placed the article's rank in the top 10 and 2) at least  $(n - 1)$  reviewers ranked the article in their individual top 10 rankings, where  $n$  is the number of eligible reviewers. Although there was a paucity of published qualitative studies, this year two of the three papers had scores that were consistent with previously highlighted qualitative papers, and therefore both of these studies were selected as highlighted articles.

## RESULTS

A total of 243 papers satisfied the search criteria, and 25 papers met the inclusion criteria.<sup>13-37</sup> The 25 articles (22 quantitative, three qualitative studies) were critically appraised by each of six reviewers. Five quantitative studies met a priori criteria as methodologically superior publications in education research, with a range of mean scores from 15.4 to 21.2 (maximum 25 points).<sup>17,21,22,31,37</sup> Two qualitative studies met our threshold for being selected as superior with scores of 20 and 20.4 (maximum 25 points).<sup>20,32</sup> The seven highest ranking quantitative and qualitative studies are presented in alphabetical order by the surname of the first author.

Breckwoldt J, Svensson J, Lingemann C, Gruber H. Does clinical teacher training always improve teaching effectiveness as opposed to no teacher training? A randomized controlled study. *BMC Med Educ* 2014;14:6.

*Background:* Although teacher training is widely employed in faculty development programs, recent expertise research suggests that the immediate effect of training may actually be decreased performance before higher levels of competence are reached. This

Table 1  
EM Education Research Scoring Metrics: Quantitative Research

Domain	Item	Item Score	Maximum Domain Score
Introduction: give 1 point for each criterion met	Appropriate description of background literature	1	3
	Clearly frame the problem	1	
	Clear objective / hypothesis	1	
Measurement: give 0 or 1 point for each criterion met			4
1. Methodology			
	Has no pretest or posttest	0	
	Has a posttest only	1	
	Has a pretest and posttest	1	
2. Groups			
	Both experimental and control group	1	
	Random assignment to groups	1	
Data collection: give 0 or 1 point for each criterion met			4
1. Institutions			
	Single institution	0	
	At least two institutions	1	
	More than two institutions	1	
2. Response rate			
	Response rate < 50% or not reported	0	
	Response rate ≥ 50%	1	
	Response rate ≥ 75%	1	
Data analysis: give 0 or 1 point for each criterion met			3
1. Appropriateness			
	Data analysis inappropriate for study design/type of data	0	
	Data analysis appropriate for study design and type of data	1	
2. Sophistication			
	Descriptive analysis only	0	
	Beyond descriptive analysis	1	
	Includes power analysis	1	
Discussion: give 1 point for each criterion met			3
	Data supports conclusion	1	
	Conclusion clearly addresses hypothesis/objective	1	
	Conclusions placed in context of literature	1	
Limitations: assign a single best score			2
	Limitations not identified accurately	0	
	Some limitations identified	1	
	Limitations well addressed	2	
Innovation of project: assign a single best score			2
	Previously described methods	0	
	New use for known assessment	1	
	New assessment methodology	2	
Relevance of project: assign a single best score			2
	Impractical to most programs	0	
	Relevant to some	1	
	Highly generalizable	2	
Clarity of writing: assign a single best score			2
	Unsatisfactory	0	
	Fair	1	
	Excellent	2	
Total			25

“U-shaped” developmental curve may be due to disruption of current routines and knowledge organization, insecurity due to a new awareness of one’s deficits, need for time for deliberate practice to integrate new strategies into practice, and other factors. This study explored the effect of clinical teacher training on teacher quality and effectiveness at an early time point after intervention.

**Methods:** This prospective study randomized 18 clinical teachers, matched for clinical and teaching experience, to receive either a 2-day teacher training session (intervention) or no intervention (control). The training session used small groups, role-play, and reflections exercise to teach topics such as role of the teacher, learner needs, providing feedback, session instruction,

defining learning objectives, activating learners, teaching skills, and teaching with patients. Both groups participated in a 12-hour EM teaching course 2 to 3 weeks after the intervention group underwent the training session. Quality and effectiveness of teaching performance was measured in both groups by 1) student performance on a structured clinical examination (SCE) and multiple-choice question test (MCQ); 2) a 30-item student assessment of teacher performance; and 3) self-reported teacher assessment of competence, self-confidence, and satisfaction with teaching on a seven-point Likert-type scale.

**Results:** Outcomes were compared between 96 students who were taught by trained teachers and 97 students who were taught by untrained teachers. Students

Table 2  
EM Education Research Scoring Metrics: Qualitative Research

Domain	Item	Item Score	Maximum Domain Score
Introduction: give 1 point for each criterion met	Appropriate description of background literature	1	3
	Clearly frame the problem	1	
	Clear objective/hypothesis	1	
Measurement: give 1 point for each criterion met	1. Methodology		3
	Appropriate for study question	1	
	2. Sampling of participants		
	Appropriate study population	1	
	Enrolled full range of cases/settings beyond convenience	1	
Data collection: give 0–1 point for each criterion met			3
1. Institutions			
	Single institution	0	
	At least two institutions	1	
	More than two institutions	1	
2. Sample size determination			
	Appropriate sample size determination	1	
Data analysis: give 1 point for each criterion met			5
	Clear, reproducible “audit trail” documenting systematic procedure for analysis	1	
	Data saturation through a systematic iterative process of analysis	1	
	Addressed contradictory responses	1	
	Incorporated validation strategies (e.g., member checking, triangulation)	1	
	Addressed reflexivity (impact of researcher’s background, position, biases on study)	1	
Discussion: give 1 point for each criterion met	Data supports conclusion	1	3
	Conclusion clearly addresses hypothesis/objective	1	
	Conclusions placed in context of literature	1	
Limitations: assign a score	Limitations not identified accurately	0	2
	Some limitations identified	1	
	Limitations well addressed	2	
Innovation of project: assign a score	Previously described methods	0	2
	New use for known assessment	1	
	New assessment methodology	2	
Relevance of project: assign a score	Impractical to most programs	0	2
	Relevant to some	1	
	Highly generalizable	2	
Clarity of writing: assign a score	Unsatisfactory	0	2
	Fair	1	
	Excellent	2	
Total			25

taught by untrained teachers performed better on two of the four domains of the SCE, but there were no significant differences in overall SCE or MCQ scores between the two groups. Overall, teaching quality was rated significantly better by students of untrained teachers ( $p = 0.05$ ). Qualitative data suggested that trained teachers felt more competence and more self-confident than untrained teachers.

*Strengths of this study:* This study addresses a gap in the medical education literature regarding the immediate effects of professional development in teaching on teacher and learner performance and is relevant to educators. The randomized, controlled design is a strength, as is the emphasis on learner performance outcomes.

*Relevance for future educational advances:* Future studies might enroll teachers at multiple sites and follow them over time to determine if after a period of adjustment training does eventually improve teacher performance.

Chan T, Bakewell F, Orlich D, Sherbino J. Conflict prevention, conflict mitigation, and manifestations of conflict during emergency department consultations. *Acad Emerg Med* 2014;21:308–13.

*Background:* Consultation, defined as a request for admission or further consulting service management of a patient, is common in emergency department (ED) visits and requires collaboration and balancing competing interests. Little is known about how conflict affects ED consultation. This qualitative study aimed to explore the theme of conflict in ED consultation, as perceived by emergency physicians (EPs) and consultants.

*Methods:* Thirty-one residents and 30 attending physicians from EM, general surgery, and internal medicine were interviewed regarding two conflict-related questions: 1) what are the impediments to a good consultation-referral encounter and 2) how do you cope or compensate for these impediments? Interviews were

recorded, transcribed, and analyzed using a grounded theory approach.

**Results:** Conflict-mitigating and -producing themes identified related to historical factors, attitudes and values, and actions. Positive themes included good reputation, good prior experiences, empathy, engagement, professional behavior, agreeing with plan of care, collaboration, meeting expectations, provision of expert care, adjusting expectations, and presence. Negative themes included bad or unknown reputation, doubt in other party's competence, being disengaged, poor communication, disagreeing with care plan, self-serving behaviors, providing inappropriate care, failing to collaborate, failing to meet expectations, absence, and external stressors. Trust was a theme that intersected all other themes, as breaches in trust can lead to conflict, and trust-building activities may mitigate or resolve conflict.

**Strengths of this study:** The theme of conflict, and how it affects ED consultations, is relevant to education and clinical practice and has not been explored in the ED setting. This qualitative study used a rigorous approach to begin to understand the phenomenon of ED consultation conflict, as perceived by EPs and consultants.

**Relevance for future educational advances:** This study highlights important themes in ED conflict, which is a threat to effective collaboration with consultants. Future studies might explore methods of measuring conflict or effective consultations, develop and test the effect of interventions designed to improve consult collaboration, or explore cultural factors that may affect conflict in the ED.

Chang T, Pham P, Sobolewski B, et al. Pediatric emergency medicine asynchronous E-learning: a multicenter randomized controlled Solomon four-group study. *Acad Emerg Med* 2014;21:912–19.

**Background:** Asynchronous e-learning methodology has become commonplace in medical education. Asynchronous learning can provide educational material in a standardized fashion to a variety of learners despite significant variability in clinical responsibilities and schedule. Short-term knowledge benefits have been shown mostly in local, single-venue studies.

**Methods:** This was a multi-institutional, randomized trial using the Solomon four-group randomization scheme. Numerous learners from levels ranging from fourth-year medical student to postgraduate year [PGY]-4 resident rotating in four U.S.-based pediatric EDs were randomized to perform a series of asynchronous learning modules or to a control group that did not perform the modules during the rotation. All four groups were posttested using a 30-question internally validated posttest; two of the four groups (one e-module and one control) were pretested. The outcomes of interest were performance on the posttest based on group assignment and improvement from pre- to posttesting. In addition, the authors analyzed performance as a function of year of training.

**Results:** A total of 256 participants fully completed the study protocol, the majority (173) being residents. A significant number of potential participants dropped out (44%), mostly due to incompleteness of learning modules.

Use of e-learning modules increased knowledge as measured by improvement from pre- to posttest ( $p < 0.001$ ) with a large effect size. When broken down by learner, EM and pediatric residents showed improvement based on module use, while family medicine residents and medical students did not. Number of modules performed correlated with posttest improvement ( $r^2 = 0.14$ ).

**Strengths of this study:** This study was multi-institutional, with large numbers of learners from a variety of specialties and learning levels. The methodological design using the Solomon four-group randomization scheme was elegant. The Solomon four-group design attempts to mitigate bias introduced by test sensitization, randomizing participants not only to "intervention" and "control" groups, but also to intervention and control groups that receive a pretest and those that do not. The curricular design of the e-learning modules was well described and thorough. The generalizability is high, given that asynchronous learning is popular and the study demonstrated the ability to disseminate curricular enhancements to multiple institutions.

**Relevance for future educational advancements:** Well-designed, multi-institutional educational research is possible, as the authors elegantly show. Common concerns about pretest sensitization can be mitigated. Participant completion of protocols remains a challenge, and efforts can be made to increase full curricular involvement.

Clarke S, Horeczko T, Carlisle M, et al. Emergency medicine resident crisis resource management ability: a simulation-based longitudinal study. *Med Educ Online* 2014;19:25771.

**Background:** The purpose of this study was to test whether a validated assessment tool for crisis resource management (CRM) skills could demonstrate EM resident growth in team-leadership skills through residency.

**Methods:** This was a single-institution, nonblinded cohort study of residents progressing through a 3-year program. The residents were individually tested annually, midyear, via a simulation using the Ottawa Crisis Resource Management Global Rating Scale. Forty-five residents were tested over a full 3-year cycle.

**Results:** First-year residents demonstrated wide variability on a scale of leadership communication skills. These skills universally improved and the variability between residents narrowed in the second year. There was no clear progression between the second and third year of residency. The findings were significant for overall performance, leadership, problem solving, and resource utilization. Situational awareness and communication skills did not show significant results. The conclusion is that this validated tool could be used to demonstrate progression of CRM skills during residency.

**Strengths of the study:** While a modest study, this paper follows a cohort of residents over 3 years with repeat measures. The design is simple, the statistics are uncomplicated, and the results are stated succinctly and correctly without being overstated. The paper convincingly demonstrates that while residents begin with a large variability in CRM skills, these skills improve and variability is reduced over the first year. The residency

cohort as a whole did not progress in these communication skills after the first year.

*Relevance for future educational advances:* This study presents a baseline and methodology for assessing CRM skills for future innovations in training. It also suggests that individual resident testing should be focused on second-year residents to identify the few outliers who do not follow the predicted growth rate of the majority. This study would need to be validated in other settings before being broadly accepted.

Maddy J, Varney S, Sessions D, et al. A comparison of simulation-based education versus lecture-based instruction for toxicology training in emergency medicine residents. *J Med Toxicol* 2014;10:364–8.

*Background:* Toxicologic emergencies are rare events that require specific intervention after appropriate identification of the inciting agent in critically ill poisoned patients. The authors hypothesized that simulation would improve EM resident medical knowledge and comfort compared to traditional lecture-based education at 3 months after training.

*Methods:* This study block-randomized EM residents to receive either simulation-based (intervention) or standard lecture (control) instruction on four toxicologic emergencies. Medical knowledge (measured by a 24-item vignette-based multiple-choice test) and self-report comfort with managing each scenario were assessed preintervention, immediately postintervention, and 3 months postintervention. Knowledge and comfort outcomes were compared between groups at every time point.

*Results:* Sixteen residents were randomized into the simulation group, and 17 were randomized into the lecture group. There were no significant differences in baseline medical knowledge or comfort scores at baseline. The lecture group had a higher immediate medical knowledge score improvement than the simulation group (5.6- vs. 3.6-point increase;  $p = 0.02$ ), but of 24 residents who completed the 3-month follow up (11 in the simulation arm, 13 in the lecture arm), the simulation group had modest but significantly higher final mean test scores (16.6 vs. 13.3;  $p = 0.009$ ). There were no differences in comfort scores immediately or at 3 months postintervention.

*Strengths of this study:* This study compared simulation to standard instruction, using a randomized design to attempt to evaluate the benefit of simulation, and measured 3-month knowledge retention in addition to immediate postintervention knowledge. Outcomes include both self-reported comfort and multiple-choice test assessment of medical knowledge.

*Relevance for future educational advances:* Determining the effect of simulation on long-term knowledge retention, compared with other didactic methods, is a topic that warrants further study. Assessing simulation performance beyond medical knowledge, with a greater number of learners, and looking at retention past 3 months are important next steps.

Pelaccia T, Tardif J, Tribby E, et al. How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Ann Emerg Med* 2014;64:575–85.

*Background:* The goal of this qualitative study was to identify how and when, during the initial patient examination, EPs generate and evaluate diagnostic hypotheses.

*Methods:* Participating expert EPs at three clinical study sites wore head-mounted video cameras to record their initial encounters with actual patients presenting with potentially life-threatening complaints. Semistructured interviews were subsequently performed, using a video cued-recall technique to facilitate each EP's recall and discussion of his or her reasoning during the patient encounter. Participants were asked to discuss when and how they generated their diagnostic hypotheses, as well as their certainty of hypotheses, and the complexity of each case. Data collection and thematic analysis took place iteratively to identify emerging themes related to the EPs' diagnostic reasoning. Interviews were recorded, transcribed, and analyzed using qualitative software. Preliminary and end coding was conducted by each of the primary investigators, and individual and overlapping matrices of themes were developed for the participants' responses.

*Results:* Fifteen EPs with a mean of 12 years of experience participated in the study. Interviews lasted a mean of 53 minutes, discussing initial patient encounters of a median of 14 minutes' duration. Of a total of 79 hypotheses generated by the physicians, 75% were generated before meeting the patient, based on triage information, or within the first 5 minutes of the patient encounter. Hypothesis generation was generally unconscious and automatic, prompted by the context of the presentation or "key words" during the patient interview. Levels of certainty were immediately ascribed to each hypothesis, and data collection through history and physical examination was used to support or refute each hypothesis. Hypotheses were rank ordered by the completion of the initial evaluation, the most likely hypothesis most commonly the one generated within the first 90 seconds of the encounter. Levels of certainty were never 100% until diagnostic results were complete.

*Strengths of the study:* This was a well-described qualitative analysis of cognitive decision-making in EPs, using the innovative technique of head-mounted video camera recording to cue recall in participants. The findings are valid and applicable to the contemporary emergence in the literature about how physicians make first-impression prioritized differential diagnoses.

*Relevance for future educational advances:* This study is relevant to the study and teaching of decision-making skills in EM. It is an advance in the use of qualitative research to further the understanding of clinical behaviors that cannot be quantitatively measured.

Wilson A, Pike G, Humbert A. Analyzing script concordance test scoring methods and items by difficulty and type. *Teach Learn Med* 2014;26:135–45.

*Background:* Script concordance tests (SCTs) have been used to assess the clinical judgment skills of examinees based on their interpretation of clinical data. Examinees' subsequent scores are assigned in comparison to those of a panel of experienced physicians. The goal of this study was to assess the relationship between six SCT scoring methods and the discriminatory properties of the test.

**Methods:** This was a retrospective analysis of the psychometric properties of six SCT scoring methods from a single institution with faculty and learners dispersed over nine sites. The data sets consisted of a problem-solving SCT (SCT-PS) administered to second- and fourth-year medical students (MS2, MS4) and an EM SCT (SCT-EM) administered to PGY 1 through 3 EM residents and EM faculty. A repeated-measures analysis of variance, performed on the SCT-PS data set, assessed whether scoring methods could discriminate between MS2 and MS4 training levels within the same population of student examinees. A multivariate analysis of variance was conducted on the SCT-EM data set to test whether significant differences between training levels (MS4, EM-PGY1, EM-PGY2, EM-PGY3, and EM faculty) could be detected for each scoring method.

**Results:** A total of 522 medical students completed the SCT-PS. Regardless of the scoring method used, the SCT-PS demonstrated moderate predictive validity and discrimination between MS2 and MS4 scores. There was a statistically significant increase in scores between MS2 and MS4 years across all levels of test item difficulty. A total of 988 MS4, 40 EM residents, and 12 EM faculty completed the SCT-EM. Again, regardless of scoring method, there was significant discrimination in test performance across levels of training and test item difficulty. Item type (diagnostic, investigational, therapeutic) scores correlated with level of training. The investigators found that a five-point Likert scale scoring method was moderately more reliable than a three-point scale.

**Strengths of the study:** This was a statistically rigorous analysis of a very large data set, allowing for comparisons of SCT-PS performance within one MS cohort across 2 years of training, as well as SCT-EM performance across multiple levels of training and clinical EM experience.

**Relevance for future educational advances:** This study demonstrates that if SCTs are used to assess clinical decision-making in EM, a five-point Likert scale may be moderately more reliable in scoring, but that regardless of scoring method, SCTs can differentiate clinical judgment skills across levels of training.

## DISCUSSION

### Trends in Medical Education Research 2014

In 2014 significantly fewer articles met our criteria for full review ( $n = 25$ ) compared with 2013 ( $n = 43$ ). One disturbing finding was the decrease in funding for research in medical education this year. Reed et al.<sup>38</sup> demonstrated that studies that were funded tended to be of higher quality when assessed on a validated tool. In 2014, only four studies (16%) were funded, three by professional societies and one by a governmental agency.<sup>20,25,32,35</sup> This represents a sharp decrease from 2013, when 32.5% reported some sort of financial support. Two of the five highlighted articles this year received funding.<sup>20,32</sup>

Over half (56%) of fully reviewed articles were published in EM journals this year, which is similar to 2013. Eight (32%) were published in medical education journals, which represents a notable increase since last year

(19%). Three (12%) appeared in the specialty journal that was the topic of the study within EM (surgery, toxicology, ultrasound). Twenty-one (84%) articles had EM authors, a slight decline from the 91% found in 2013. In five of these articles, the EM author collaborated with a specialist in medical education.

Observational studies predominated in 2014, with 13 (52%) employing this method.<sup>13,15,16,18,22,24,26,27,33,35-37</sup> Three articles (12%) featured qualitative methods, and unlike previous years, all appeared in EM journals.<sup>20,28,32</sup> Two of these studies are in the highlighted group.<sup>20,32</sup> In general, qualitative methods enable the research team to identify themes that can later serve as the basis for hypothesis-driven research. It is incumbent on education researchers in EM to understand the basic tenets of qualitative research and its role in laying the groundwork for future study. Experimental or quasi-experimental research design accounted for nine (36%) of the studies, a slight increase over 2013 (33%).<sup>14,17,19,21,25,29-31,34</sup> Outcomes derived from research measuring quantitative variables related to learning have a varying effects. In 1959, Kirkpatrick and Kirkpatrick<sup>39</sup> introduced a four-level training scheme that demonstrated progressive mastery of learned material, which has been updated over time. Beginning with learner reactions, attitudes, or self-perceived levels of comfort, Level 1 outcomes are an important foundation for future research, but are not generalizable or predictive of actual performance. Level 2 outcomes demonstrate learning, in that there is some documented improvement in mastery of the subject matter in the classroom or laboratory environment. Of significant interest to educators are Kirkpatrick Level 3 outcomes, in which learner behavior has become modified as a result of the educational intervention. Success in this level demonstrates that the learners are actually applying what they have learned to their daily practice. Seven (28%) of this year's articles fall into this category, including two of the featured articles.<sup>13,17,22,25,26,29,34</sup> As a specialty, it would be desirable for our research to demonstrate Level 4 outcomes. In this ideal situation, an intervention would not only create the desired behavior in the trainees, but this behavior would also result in improved outcomes in the targeted areas when measured independently.

Authors studied physicians in training and in practice, but no studies focused on patients or other groups this year. Medical student education was the central theme in five (20%) of studies,<sup>18,28,30,34,36</sup> while residents were the target of study in 12 (48%) studies,<sup>13-15,20,22-27,31,33</sup> and faculty were the primary group in one (4%).<sup>32</sup> Many studies targeted multiple learner groups: medical students and residents;<sup>21</sup> medical students, residents, and faculty;<sup>16,37</sup> medical students and faculty;<sup>17</sup> and residents and faculty.<sup>19,20,29,35</sup>

Four studies were carried out at two or more institutions.<sup>13,20,21,27</sup> This is very helpful to a broad base of educators who may wish to implement a similar intervention into their own program, as positive results indicate that the intervention itself and not just the particular setting and instructor may be responsible for improved learning outcomes. As a community, EM researchers can collaborate to enhance learning

**Table 3**  
Trends for the Reviewed Medical Education Research Articles of 2014

Variable	All Publications (n = 25)	Highlighted (n = 7)
Funding	4	2
Learner groups*		
Medical students	9	3
Residents	18	5
Faculty	8	4
Study methodology*		
Observational	13	2
Experimental	9	3
Qualitative	3	2
Topics of study*		
Technology	12	3
Simulation	6	2
Clinical reasoning	2	2
Teaching/learning methods	9	4
Assessment	8	2
Next accreditation system	5	1

\*It is possible to exceed the total N = 25 or N = 7 due to multiple learner categories or study topics.

across the specialty if the results of studies are more generalizable.

Topics of study in 2014 followed the ongoing trend of employing technology, with a 48% prevalence,<sup>13–15,19,22–24,30–33,35</sup> of which half were based on simulation;<sup>13,15,22,23,30,31</sup> both figures are similar to 2013. Other uses of technology included ultrasound,<sup>13,19</sup> social media,<sup>24</sup> video,<sup>32,33</sup> and audience response systems.<sup>35</sup> However, in contrast to 2013 (16%), only one study (4%) centered on pediatric EM.<sup>15</sup> Other topics in 2014 were productivity in the ED (8%)<sup>16,25</sup> and ethics (4%).<sup>28</sup> Half of the 25 articles focused on teaching or learning methods,<sup>13,14,17,18,21,29,31,32,34–37</sup> and in keeping with last year's emergence of clinical reasoning and decision-making, three were centered on this theme.<sup>32,34,37</sup> Learner assessment and competence were again important topics in 2014, with eight articles (32%).<sup>18,22–24,26,27,36,37</sup> It is possible that these studies are being prompted by the implementation of the Next Accreditation System (NAS).<sup>40</sup> Milestones for the progressive development of trainees are featured in this system, and the need to teach, observe, and assess learners is a cornerstone of residency certification in all specialties. The fact that EM educators and researchers are studying elements that are required in NAS shows a healthy approach to their implementation. Documentation of efficacy of programs is not only useful for accreditation, but can serve as a template for all EM educators whose needs are similar. A summary of the trends of articles published in 2014 is provided in Table 3.

## LIMITATIONS

Despite extensive searches using previously developed and implemented strategies of the English language literature for all publications that met review criteria,

these searches may have erroneously omitted high-quality studies. The exclusion criteria used to cull articles may be considered overly rigorous. For example, per the search criteria revision in 2012, single-site survey studies were not included in the review process. This year we also excluded studies that examined outcomes limited to an expected learning effect without a comparison group. This likely resulted in the lower numbers of papers selected for scoring, but was implemented to exclude studies with a limited ability to affect EM educational practice.

The rating metrics that have been used in this review are adapted from accepted literature on education research review and assessment and include broadly accepted metrics for experimental studies, as well as the unique metrics associated with qualitative research design. However, the scoring methods have not been externally validated. This fact may contribute to a ranking cut point of quality articles that may be too stringent.

Given that the goal of this continuing series is to review high-quality education research pertinent to the teaching of EM, the sequential adjustments to the serial article selection process and rating methods have set the rating and ranking bars higher, continuing to promote the best examples of medical education research in EM.

## CONCLUSIONS

This critical appraisal of the 2014 EM education research literature highlights quality publications and recent trends in the field. The five quantitative and two qualitative studies featured represent methodologically superior research published in 2014. Each contributes to the expanding field of education research, while addressing the methods to control, justify, or minimize the limitations that are inherent to this focus. These highlighted studies can serve as exemplary models for EM educators interested in conducting high-quality, methodologically sound education research.

The authors acknowledge Michelle Lin, MD, and Gloria J Kuhn, DO, who helped develop the initial study concept and critical appraisal procedure for this work.

## References

1. LaMantia J, Deiorio NM, Yarris LM. Executive summary: education research in emergency medicine opportunities, challenges, and strategies for success. *Acad Emerg Med* 2012;19:1319–22.
2. Love JN, Coates WC, Santen SA, Hobgood CD, Mavis BE, Farrell SE. The MERC at CORD scholars program in medical education research: a novel faculty development opportunity for emergency physicians. *Acad Emerg Med* 2009;16(Suppl 2):S37–41.
3. Society for Academic Emergency Medicine. Education Research Grant. Available at: <http://www.saem.org/saem-foundation/grants/what-we-fund/education-research-grant>. Accessed Aug 31, 2015.
4. Council of Emergency Medicine Residency Directors and Emergency Medicine Foundation. CORD-EMF

- Education Research Grant. Available at: <http://www.cordem.org/i4a/pages/index.cfm?pageid=3356>. Accessed Aug 31, 2015.
5. Fisher J, Lin M, Coates WC, et al. Critical appraisal of emergency medicine educational research: the best publications of 2011. *Acad Emerg Med* 2013;20:200–8.
  6. Boyer E. *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching, 1990. Available at: <http://www.hadinur.com/paper/BoyerScholarshipReconsidered.pdf>. Accessed Aug 31, 2015.
  7. Farrell SE, Coates WC, Khun GJ, Fisher J, Shayne P, Lin M. Highlights in emergency medicine medical education research: 2008. *Acad Emerg Med* 2009;16:1318–24.
  8. Kuhn GJ, Shayne P, Coates WC, et al. Critical appraisal of emergency medicine educational research: the best publications of 2009. *Acad Emerg Med* 2010;17(Suppl 2):S16–25.
  9. Shayne P, Coates WC, Farrell SE, et al. Critical appraisal of emergency medicine educational research: the best publications of 2010. *Acad Emerg Med* 2011;18:1081–9.
  10. Cote L, Turgeon J. Appraising qualitative research articles in medicine and medical education. *Med Teach* 2005;27:71–5.
  11. Daley J, Willis K, Small R, et al. A hierarchy of evidence for assessing qualitative health research. *J Clin Epidemiol* 2007;60:43–9.
  12. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245–51.
  13. Adhikari S, Raio C, Morrison D, et al. Do emergency ultrasound fellowship programs impact emergency medicine residents' ultrasound education? *J Ultrasound Med* 2014;33:999–1004.
  14. Auten J, Ross E, French M, et al. Low-fidelity hybrid sexual assault simulation training's effect on the comfort and competency of resident physicians. *J Emerg Med* 2015;48:344–50.
  15. Bank I, Snell L, Bhanji F. Pediatric crisis resource management training improves emergency medicine trainees: perceived ability to manage emergencies and ability to identify teamwork errors. *Pediatr Emerg Care* 2014;30:879–83.
  16. Bhat R, Dubin J, Maloy K. Impact of learners on emergency medicine attending physician productivity. *West J Emerg Med* 2014;15:41–4.
  17. Breckwoldt J, Svensson J, Lingemann C, Gruber H. Does clinical teacher training always improve teaching effectiveness as opposed to no teacher training? A randomized controlled study. *BMC Med Educ* 2014;14:6.
  18. Carr S, Celenza A, Puddey I, Lake F. Relationships between academic performance of medical students and their workplace performance as junior doctors. *BMC Med Educ* 2014;14:157.
  19. Carrie C, Biais M, Lafitte S, Grenier N, Revel P, Janvier G. Goal-directed ultrasound in emergency medicine: evaluation of a specific training program using an ultrasonic stethoscope. *Eur J Emerg Med* 2014; epub ahead of print.
  20. Chan T, Bakewell F, Orlich D, Sherbino J. Conflict prevention, conflict mitigation, and manifestations of conflict during emergency department consultations. *Acad Emerg Med* 2014;21:308–13.
  21. Chang T, Pham P, Sobolewski B, et al. Pediatric emergency medicine asynchronous E-learning: a multicenter randomized controlled Solomon four-group study. *Acad Emerg Med* 2014;21:912–9.
  22. Clarke S, Horeczko T, Carlisle M, et al. Emergency medicine resident crisis resource management ability: a simulation-based longitudinal study. *Med Educ Online* 2014;19:25771.
  23. Clarke S, Horeczko T, Cotton D, Bair A. Heart rate, anxiety and performance of residents during a simulated critical clinical encounter: a pilot study. *BMC Med Educ* 2014;14:153.
  24. Desai B. A novel use of Twitter to provide feedback and evaluations. *Clin Teach* 2014;11:141–5.
  25. Dubosh N, Carney D, Fisher J, Tibbles C. Implementation of an emergency department sign-out checklist improves transfer of information at shift change. *J Emerg Med* 2014;47:580–5.
  26. Hauff S, Hopson L, Losman E, et al. Programmatic assessment of level 1 Milestones in incoming interns. *Acad Emerg Med* 2014;21:694–8.
  27. Hopson L, Burkhardt J, Stansfield R, Vohra T, Turner-Lawrence D, Losman E. The multiple mini-interview for emergency medicine resident selection. *J Emerg Med* 2014;46:537–43.
  28. House J, Theyyunni N, Barnosky A, et al. Understanding ethical dilemmas in the emergency department: views from medical students' essays. *J Emerg Med* 2015;48:492–8.
  29. Hughes K, Benenson R, Krichen A, Clancy K, Ryan J, Hammond C. A crew resource management program tailored to trauma resuscitation improves team behavior and communication. *J Am Coll Surg* 2014;219:545–51.
  30. Karakuş A, Duran L, Yavuz Y, Altintop L, Calişkan F. Computer-based simulation training in emergency medicine designed in the light of malpractice cases. *BMC Med Educ* 2014;14:155.
  31. Maddry J, Varney S, Sessions D, et al. A comparison of simulation-based education versus lecture-based instruction for toxicology training in emergency medicine residents. *J Med Toxicol* 2014;10:364–8.
  32. Pelaccia T, Tardif J, Tribby E, et al. How and when do expert emergency physicians generate and evaluate diagnostic hypotheses? A qualitative study using head-mounted video cued-recall interviews. *Ann Emerg Med* 2014;64:575–85.
  33. Sakles J, Mosier J, Patanwala A, Dicken J. Learning curves for direct laryngoscopy and GlideScope® video laryngoscopy in an emergency medicine residency. *West J Emerg Med* 2014;15:930–7.
  34. Sherbino J, Kulasegaram K, Howey E, Norman G. Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *Can J Emerg Med* 2014;16:34–40.

35. Stoneking L, Grall K, Min A, Dreifuss B, Spear Ellinwood K. Role of an audience response system in didactic attendance and assessment. *J Grad Med Educ* 2014;6:335–7.
36. Wallenstein J, Ander D. Objective structured clinical examinations provide valid clinical skills assessment in emergency medicine education. *West J Emerg Med* 2015;16:121–6.
37. Wilson A, Pike G, Humbert A. Analyzing script concordance test scoring methods and items by difficulty and type. *Teach Learn Med* 2014;26:135–45.
38. Reed D, Cook D, Beckman T, Levine R, Kern D, Wright S. Association between funding and quality of published medical education research. *JAMA* 2007;298:1002–9.
39. Kirkpatrick D, Kirkpatrick J. *Evaluating Training Programs*. San Francisco, CA: Berrett-Koehler, 2006.
40. Nasca T, Philibert I, Brigham T, Flynn T. The next GME accreditation system—rationale and benefits. *N Engl J Med* 2012;366:1051–6.