

Flipping the Classroom in Graduate Medical Education: A Systematic Review

Andrew M. King, MD
Michael Gottlieb, MD
Jennifer Mitzman, MD

Tina Dulani, MD
Stephanie J. Schulte, MLIS
David P. Way, MEd

ABSTRACT

Background Flipped classroom (FC) instruction has become increasingly common in graduate medical education (GME).

Objective The purpose of this study was to profile the use of FC in the GME setting and assess the current status of research quality.

Methods We conducted a systematic literature search of major health and social science databases from July 2017 to July 2018. Articles were screened to ensure they described use of the FC method in an Accreditation Council for Graduate Medical Education–accredited residency program and included research outcomes. Resulting articles were analyzed, described, and evaluated for research quality using the Kirkpatrick framework and the Medical Education Research Study Quality Instrument (MERSQI).

Results Twenty-two articles were identified, all of which were recently published. Five were only indirectly related to FC methods. Most studies reported Kirkpatrick-level outcomes. Studies involving resident learner opinions were generally positive. Pre-posttest studies resulted in large positive improvements in knowledge or skills attainment. Control group study results ranged from large positive (1.56) to negative effects (−0.51). Average MERSQI scores of 12.1 (range, 8.5–15.5) were comparable to GME research norms.

Conclusions Varying methods for implementing and studying the FC in GME has led to variable results. While residents expressed a positive attitude toward FC learning, shortcomings were reported. Approximately half of the studies comparing the flipped to the traditional classroom reported better achievement under the FC design. As indicated by the MERSQI score, studies captured by this review, on average, were as rigorous as typical research on residency education.

Introduction

Medical education has been shifting from traditional, lecture-based teaching to approaches that promote higher-order thinking and active learning.¹ One particular approach, the flipped classroom (FC), has become increasingly popular in higher education, including medical education.² In this educational pedagogy, foundational content materials are studied by the learner independently through preclassroom activities, such as reading an article or textbook chapter, watching a multimedia presentation, or listening to a lecture in advance of classroom time. The classroom is reserved for applying foundational content knowledge in small group discussions involving clinical cases or more generic problem-solving. Conversely, in the traditional classroom, foundational content material is transmitted to learners through lectures that require review and reinforcement through study after class.^{3,4}

Advances in technology and the search for more effective approaches to teaching seem to be driving the shift to FC.⁵ Active, self-directed learning—a necessary component of the FC learning model—is consistent with the needs of resident learners.⁴ Additionally, this model is consistent with social, behavioral, and constructivist learning theories.^{6,7} Group collaboration encourages modeling, scaffolding, and feedback that engage learners and facilitate the integration of new knowledge with old knowledge.⁸ Compared to the traditional lecture, the FC promotes higher levels of cognitive processing as defined by Bloom's taxonomy.^{8–10} As a result, increasing numbers of educators have adopted the FC across a variety of educational settings.^{11–19}

Literature on the FC has proliferated rapidly across health sciences education since its inception in 2007.²⁰ Numerous articles originated from pharmacy, nursing, or veterinary medicine education programs.^{3,7,12,21–29} Both a recent systematic review and a meta-analysis covering the FC in medical education yielded very few articles on graduate medical education (GME).^{3,30}

The GME setting is more challenging than the undergraduate medical education (UME) setting. Resident learners may be less motivated by grades

DOI: <http://dx.doi.org/10.4300/JGME-D-18-00350.2>

Editor's Note: The online version of this article contains the full search strategies for all included databases and Medical Education Research Study Quality Instrument (MERSQI) items with corresponding article scoring.

and more motivated by learning that facilitates patient care. Additionally, the GME learner's primary role is care provider while the UME learner's primary role is student. Given the marked differences in learning environments between UME and GME, and the FC's increasing popularity with residency programs,¹⁹ a review specific to the effectiveness of the FC method in GME was needed. The goal of this systematic review is to profile the use of and assess the quality of the research literature on FC methods used in GME.

Methods

This study conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews and was performed in accordance with best practice guidelines.³¹

A health sciences librarian (S.J.S.) performed comprehensive searches of multiple databases, including PubMed, CINAHL Plus, Embase, Web of Science Core Collection, and ERIC, on July 27, 2017. Search alerts were used to gather new records through July 27, 2018. Major search terms and strategies are provided as online supplemental material. To identify additional manuscripts, we also searched MedED-PORTAL and reviewed bibliographies of included studies.

Inclusion and Exclusion Criteria

All articles published through July 2018 involving the use of FC in Accreditation Council for Graduate Medical Education (ACGME)-accredited residency programs were screened. We included articles that pertained to residents in any year of training from any medical specialty. Other inclusion criteria involved peer-reviewed, full-text articles that described some form of research or evaluation. We excluded articles that involved FC in non-medical education settings and at other levels of medical education, including UME, continuing medical education, or fellowship. We also excluded published abstracts. We did include articles that involved research or evaluation about preclassroom learning activities associated with FC. These activities are designed to introduce new knowledge to prepare learners for the classroom session, where they will apply the new knowledge from presession activities.

Two authors (A.M.K. and M.G.) independently screened titles and abstracts of retrieved articles, and subsequently selected articles for full-text review. Discrepancies were resolved by a third reviewer (J.M.). Two authors (J.M. and T.D.) independently reviewed the full-text articles and made the determination for inclusion in the final quality review. A third

author (D.P.W.) adjudicated discrepancies at this stage.

Data Extraction

We extracted the following variables for each study: authors, publication year, medical specialty, level of trainees, education topic, FC method, specific intervention, number of FC participants, study design, and study outcomes. Research design terms used for classifying study design were derived from Campbell and Stanley.³² Originally, 2 authors (J.M. and T.D.) independently performed data extraction using a worksheet to guide the process. Differences were adjudicated by a third reviewer (M.G.). Data from articles added later in the process were extracted by 2 other authors (A.M.K. and D.P.W.).

Quality Assessment

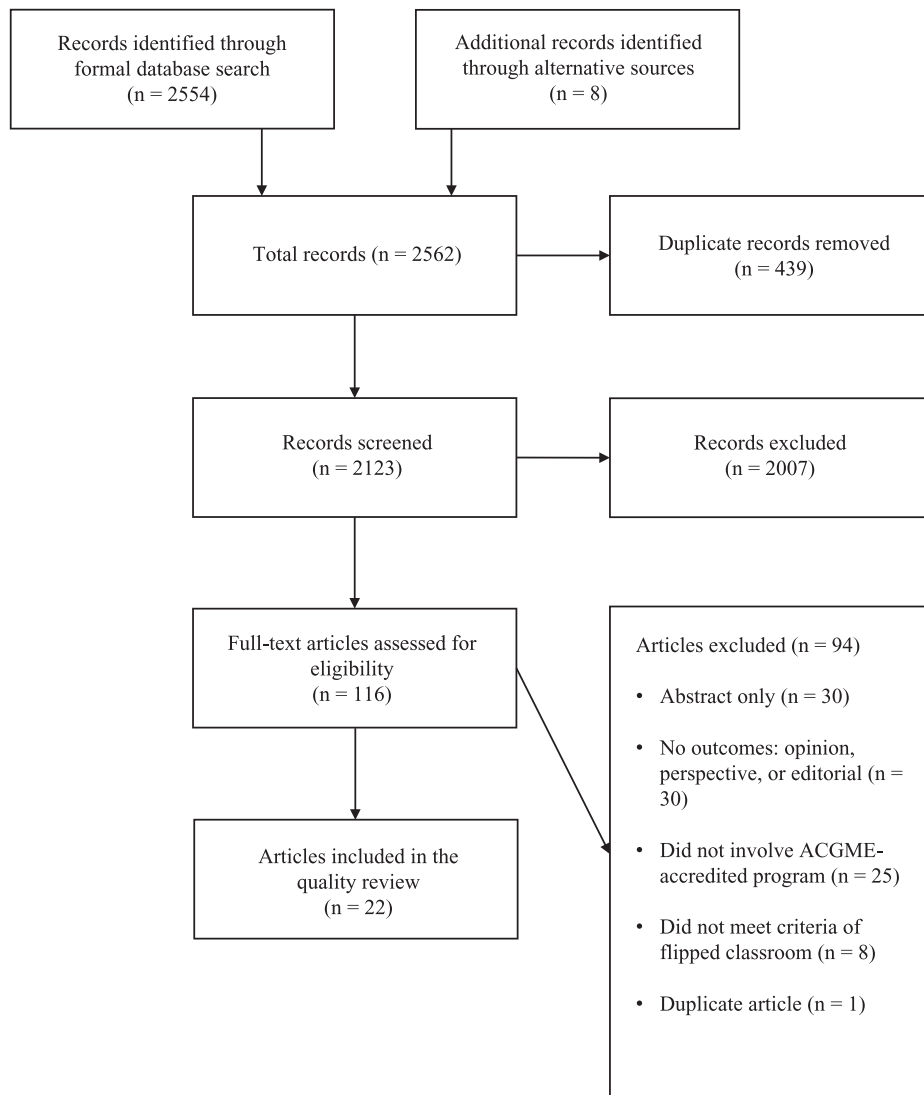
We used the modified Kirkpatrick framework to classify study outcomes of educational interventions according to impact level.³³ The modified Kirkpatrick classification levels are 1, measures of learner perceptions; 2a, self-reported changes in learner opinions; 2b, changes in knowledge or skills; 3, changes in learner behavior; 4a, change in professional practice; and 4b, change in patient's condition. When not provided, we also attempted to extract information needed to calculate associated Cohen's *d* effect sizes for each outcome.^{34,35}

Finally, we used the Medical Education Research Study Quality Instrument (MERSQI) to assess the quality of selected studies.^{36,37} This 10-item scale provides a measure of methodological quality across 6 domains: study design, sampling, type of data, validity evidence, data analysis, and outcomes. Total MERSQI scores range from 2 (low-quality research) to 18 (high-quality research). Two authors (A.M.K. and J.M.) independently scored each article. An educational researcher (D.P.W.) adjudicated discrepancies in MERSQI scoring and Kirkpatrick level assignment.

Results

Search Strategy

The FIGURE illustrates the systematic review process. The initial search yielded 2562 articles. After removing duplicates, 2123 studies were screened using titles and abstracts. Articles were excluded based on criteria outlined earlier, which resulted in 116 articles for full-text review. Of these 116, an additional 94 were excluded because they did not meet criteria. The final list of articles that met criteria for quality review included 22 articles.^{8,38-58}



FIGURE

PRISMA Diagram Illustrating Selection and Review Process of Articles Related to Flipped Classroom in Graduate Medical Education

Education Content

The 22 studies included in the final analysis are presented in TABLES 1 and 2. All were published in the past 5 years (2014–2018), most (86%) in the past 3 years. Across all studies, approximately 985 learners were involved with a flipped classroom intervention. Thirteen medical specialties were represented, including anesthesiology,^{46,49,50,57} emergency medicine,^{8,38,44,52,58} internal medicine,³⁹ internal medicine–pediatrics,⁴³ neurology,⁴⁷ neurological surgery,⁴¹ obstetrics and gynecology,^{55,56} pathology,⁴² pediatrics,^{40,43,51} psychiatry,⁴⁵ radiology,⁵³ thoracic surgery,⁴⁸ and surgical intensive care.⁵⁴ All study designs were quantitative; however, a few gathered feedback through open-ended survey items.

Authors of reviewed studies offered numerous reasons for incorporating the FC into their residency programs. Some suggested that using the FC was a solution to scheduling issues—either saving time in already saturated didactic schedules or providing schedule flexibility.^{38,39,41,48,52,53,56} Others suggested that the FC method was selected to improve didactic instruction, motivate or engage learners, promote active learning, or prepare for more advanced levels of content material.^{38–42,44,46,47,49,54}

Residency programs adopted the FC method in 1 of 3 ways. Most commonly, authors described “occasional” use of the FC for a single lesson, usually to emphasize an important topic such as quality improvement,^{39,43} resident-as-teacher skills,⁴⁰ using clinical guidelines,⁵¹ echocardiography or

electro-encephalograph interpretation,^{47,54,57} or pediatric advanced life-support.⁵² Other programs have replaced their entire didactic curriculum with FC sessions.^{41,44,48,50,55} Three articles describe the use of FC for portions of their didactic curriculum: intern orientation,^{38,49,53} reinforcement of important basic science principles,^{42,45,46} and weekly didactics during 1- or 2-month rotations.^{39,47,54,55} As proof-of-concept projects, Haspel and Lockhart designed FC instruction for delivery at national conferences.^{42,45}

Concerns about the FC surfaced in some of the literature, particularly around learner compliance with self-directed learning. Rose and colleagues discovered that residents did not accurately report time spent viewing video lectures in preparation for FC activities.⁵² Residents in the Young et al study suggested that time for preclass preparation was a weakness of the FC format.⁵⁸ Without adequate preclass preparation, learners are unable to effectively participate in applying content knowledge during the in-class meeting. Consequently, several studies involved the production of innovative methods for delivering preclass content such as podcast lectures, interactive electronic modules, or multimedia textbooks. These materials were designed specifically to persuade learners into completing the self-directed learning required for constructive participation during in-class activities.^{41,45,47,49,50,52,57}

Study Outcomes

Five of the selected articles focused more on preclass methods and less about the FC method itself. Lockhart and colleagues assessed the implementation of a small, private, online course.⁴⁵ Moeller et al and Vasilopoulos et al studied the effectiveness of podcast lectures.^{47,57} Rose and colleagues studied the impact of embedding multiple-choice questions into instructional videos,⁵² and Ortega et al evaluated a multimedia, interactive textbook.⁵⁰ Since none of these studies reported outcomes related to both preclass and in-class FC methods, they were not subject to Kirkpatrick classification.

Level 1 Outcomes–Perception: Most of the 7 studies that evaluated Kirkpatrick Level 1 outcomes were not designed for control group comparisons or change over time.^{41–44,48,51,58} Accordingly, effect sizes were not appropriate for Level 1 articles. An exception was the article by King and colleagues that compared ratings of curriculum components before and after the program switched from lecture-based to FC methods.⁴⁴ Residents rated all components higher under the FC model, but only 2 (adult and pediatric case

conferences) were deemed significantly higher in quality and value, which yielded large effect sizes ($d = 1.19$ and 1.10 , respectively).⁴⁴

Level 2a Outcomes–Change in Opinion: Almost a third of the articles reported changes in learner opinions about the FC, the content taught, or the learners themselves.^{38–40,43,46,49,53,54} Changes in opinion about the FC method were observed in a control group study and a pre-post study, both yielding large effect sizes of 1.1 and 0.95, respectively.^{39,46} Residents improved their opinions about teaching in a pre-post study of resident-as-teacher skills, which also yielded a large effect (0.95).⁴⁰ Studies that reported changes in self-efficacy, confidence, comfort, or anxiety yielded smaller effect sizes, ranging from 0.32–0.84.^{39,49} An exception was the Bonnes resident-as-teacher study, which generated a large 1.68 effect size for change in self-assessment of resident teaching skills.³⁹ Effect sizes could not be calculated for the remaining articles reporting Level 2a outcomes.^{38,43,53,54}

Level 2b Outcomes–Change in Knowledge or Skills: A third of the articles reported changes in knowledge or skills that came from controlled studies involving either parallel or historic control groups.^{8,39,41,44,46,51,55,56} Effect sizes for these studies ranged from large (1.56) to negative (–0.51). An additional 5 articles reported changes in knowledge or skills that came from pre-posttest studies, 2 of which had associated effect size that were either large (0.81)⁴⁰ or very large (2.73).³⁸ Effect sizes could not be calculated for the other 3 articles.^{43,53,54}

Level 3 Outcomes–Change in Behaviors: Only 2 studies reported changes in behaviors: one involving increases in self-directed learning outside of class,⁴⁴ and the other reporting inflated, self-reported, preclass preparation times when compared to actual preparation times.⁵² The effect size for the first was large (0.85),⁴⁴ but the effect size for the second could not be calculated.⁵²

Study Quality

Cook and Reed provided a table of normative data for interpreting MERSQI scores.³⁶ A mean score of 10 (range, 5–16) was generated through a review of 210 articles covering medical education.³⁶ For the specific topic of residency education, the mean MERSQI of 12.9 (range, 9–15.6) was generated through a review of 97 articles. This is considerably higher than the mean MERSQI score of 10.9 we observed for the 22 articles reviewed (range, 4–15.5, $SD = 3.08$). By dropping 4 pilot or proof-of-concept studies, which had only Level 1 outcomes from our analysis,^{45,47,50,58} the average MERSQI score would

TABLE 1

Flipped Classroom (FC) Methodology in Graduate Medical Education

First Author, y	Level of Participants	Specialty	Topic	Type of FC	How FC Is Implemented
Barrie et al, ³⁸ 2018	PGY-1	Emergency medicine	Core medical knowledge for Level 1 Milestone (MK)	Traditional small group FC	Used FC to replace lectures during intern orientation
Bonnes et al, ³⁹ 2017	PGY-1–PGY-3	Internal medicine	Quality improvement	Project-based learning with traditional FC	Used FC to replace lectures during a 1-month rotation
Chokshi et al, ⁴⁰ 2017	PGY-2	Pediatrics	Resident-as-teacher	Team-based learning	Use of FC for four 1-hour workshops delivered in 1 day
Girgis et al, ⁴¹ 2018	PGY-1–PGY-7	Neurological surgery	Core neurological surgery	Just-in-time learning	Used FC with assigned questions by level of training to replace lectures for entire didactic curriculum
Haspel et al, ⁴² 2016	PGY-1–PGY-3	Pathology	Molecular genomic pathology; next generation genomic sequencing	Team-based learning	FC with TBL used with a 4-part workshop on molecular genomics delivered at pathology conferences
Keefer et al, ⁴³ 2016	PGY-1	Pediatrics/internal medicine–pediatrics	Quality improvement	Project-based learning with traditional FC	Used FC for a 4-session quality improvement curriculum
King et al, ⁴⁴ 2018	PGY-1–PGY-3	Emergency medicine	Core curriculum content for emergency medicine residency	Traditional small group FC	Used FC to replace lectures for entire didactic curriculum
Lockhart et al, ⁴⁵ 2017	PGY-2	Psychiatry	Neuroscience of schizophrenia	Traditional small group FC	Use of a small private online course to prepare conference attendees for an FC experience
Martinelli et al, ⁴⁶ 2017	PGY-2	Anesthesiology	Pharmacology portion of the Anesthesia Board content outline	Traditional small group FC with case discussions and audience response system questions	Study comparing FC to lecture for 4-consecutive weekly sessions across 8 residency programs
Moeller et al, ⁴⁷ 2017	PGY-2–PGY-4	Neurology	EEG subcompetencies from the ACGME milestones	FC with EEG interpretation sessions	Evaluation of video-based lectures during 1-month clinical neurophysiology rotation
Mokadam et al, ⁴⁸ 2016	All levels	Thoracic surgery	Core thoracic surgery curriculum	Case-based learning	Used FC and weekly quizzes to replace lectures for entire didactic curriculum

be more comparable to the normative data reported by Cook and Reed ($M = 12.1$; range, 8.5–15.5; $SD = 1.93$).³⁶

Looking at individual MERSQI items suggests that research on the FC in GME is still in its infancy. The literature review included numerous articles that were considered pilot or proof-of-concept studies. Some specifically targeted investigations into developing content delivery methods for preclass preparation

such as video lectures,^{43,46} podcasts,^{41,51} or a multimedia interactive textbook.⁴⁴ Only half of the articles reviewed used experimental or quasi-experimental designs with control groups (either a control arm or historical controls).^{8,39,41,44,46,48,49,51,55,56} The other half were pre-experimental, involving only 1 group pre-posttest or posttest-only designs.^{38,40,42,43,45,47,49,52–54,57,58} Only 5 studies involved cross-institutional collaborations.^{8,36,39,46,47}

TABLE 1
Continued.

Olsen et al, ⁴⁹ 2018	PGY-1–PGY-2	Anesthesiology	Essential skills for beginning residents in anesthesiology	Preparation work followed by 3-day simulation boot camp	Assigned pre-session instructional videos as part of an intern boot camp for teaching ACLS skills
Ortega et al, ⁵⁰ 2017	All levels	Anesthesiology	Core anesthesia curriculum fundamentals	Problem-based and case-based discussions	Adoption of an interactive multimedia-enhanced textbook with FC to replace weekly lectures
Peterson et al, ⁵¹ 2017	PGY-1	Pediatrics	How to apply clinical guidelines	“Modified” FC in which content knowledge material is read in class	Pilot study of three 30-minute modified FC sessions on clinical practice guidelines
Riddell et al, ⁸ 2017	PGY-1–PGY-4	Emergency medicine	Acute low back pain and acute headache	Traditional small group FC	Crossover study comparing FC to lecture for 2 topics
Rose et al, ⁵² 2016	PGY-3	Emergency medicine	Advanced pediatric life support topics	Preparation materials followed by simulations and procedures lab	Assigned precourse video lectures to prepare residents for PEM simulation and procedures boot camp
Sajedi et al, ⁵³ 2018	PGY-1	Radiology	Preparing interns for overnight call	Traditional small group FC with in-class activities dedicated to correcting knowledge gaps observed in pretest	Used FC to replace noon conference lectures
Tainter et al, ⁵⁴ 2017	PGY-2–PGY-4	Resident rotators in SICU	Point-of-care echocardiography	Preparation materials followed by interactive practice sessions	Used FC on SICU rotation over 4 sessions
Urban et al, ⁵⁵ 2016	PGY-1–PGY-4	Obstetrics and gynecology	Core gynecology-oncology curriculum	Traditional small group FC with identified learning points incorporated into case discussions on rounds	Used FC to replace lectures on gynecology-oncology rotation (1 hour per week for 8 weeks)
Valente et al, ⁵⁶ 2018	PGY-1–PGY-2	Obstetrics and gynecology	Core content for obstetrics and gynecology in-training examination	Modified team-based learning without the group assessment test	Used FC to replace lectures for entire didactic curriculum
Vasilopoulos et al, ⁵⁷ 2015	Residents and medical students	Anesthesiology	EEG interpretation	Flipped instruction involving self-study using podcasts followed by guided instruction	Evaluation of podcasting as a tool, which while not specified, used an FC model of instruction
Young et al, ⁵⁸ 2014	PGY-1–PGY-3	Emergency medicine	Syncope, pediatric gastrointestinal presentations	Traditional small group FC	Pilot study of 2 FC sessions offered during didactic schedule

Abbreviations: PGY, postgraduate year; MK, medical knowledge; TBL, team-based learning; ACLS, advanced cardiovascular life support; PEM, pediatric emergency medicine; SICU, surgical intensive care unit; EEG, electroencephalogram.

Note: Dark shading indicates studies that focus predominantly on pre-class methods rather than the FC method.

TABLE 2

Summary of Study Outcomes and Analysis

First Author, y	Study Design	Kirkpatrick Levels and Study Outcomes With Associated Effect Sizes ^a	MERSQI Score (% of 18)
Barrie et al, ³⁸ 2018	One-group pretest-posttest design (n = 12)	2a. Participants felt more proficient with the core content of the curriculum ($d = \text{NA}$). 2b. 1. Average gains on comprehensive knowledge test of 12.6 percentage points from pre- to posttest ($d = 2.73$ [1.6–3.8]).	12.5 (69)
Bonnes et al, ³⁹ 2017	Pretest-posttest control group design (FC = 95, control = 48)	2a. Residents in FC showed significant increase in preference for it compared to control ($d = 1.1$ [0.5–1.7]). 2a. Perception about online learning modules, in-class projects, and working in teams increased significantly ($d = 0.35$ [–0.15–0.85]; $d = 0.32$ [–0.18–0.81]; $d = 0.32$ [–0.18–0.82]). 2b. Knowledge scores on QIKAT were significantly better for FC group compared to control group ($d = 1.02$ [0.5–1.5]).	13.5 (75)
Chokshi et al, ⁴⁰ 2017	One-group pretest-posttest design (n = 29)	2a. Residents showed improved attitudes toward teaching after the training session ($d = 0.95$ [0.40–1.49]). 2a. Residents rated themselves as more effective teachers after the session ($d = 1.68$ [1.2–2.28]). 2b. Participants performed significantly better on observed structured teaching examinations: Teaching a skill ($d = 1.10$ [0.5–1.7]); Giving feedback ($d = 0.81$ [0.2–1.4]); Orienting a learner ($d = 1.06$ [0.4–1.7]).	13 (72)
Girgis et al, ⁴¹ 2018	Static group comparison (n = 12)	1. Participants expressed a preference for the FC ($d = \text{NA}$). 2b. Performance on board examination improved significantly ($d = 1.56$ [0.6–2.40]).	11 (61)
Haspel et al, ⁴² 2016	Postevaluation survey only (n = 62)	1. Participants evaluated the TBL workshop positively ($d = \text{NA}$).	8.5 (47)
Keefer et al, ⁴³ 2016	One-group pretest-posttest design (n = 54)	1. Participants were more comfortable about conducting QI projects ($d = \text{NA}$). 2b. Scores on content knowledge tests were higher on posttest than on pretest ($d = \text{NA}$).	9 (50)
King et al, ⁴⁴ 2018	Static group comparison (FC = 101, control = 86)	1. Residents in FC rated most components of new curriculum higher in value and quality than those in lecture. Two components (adult and pediatric case conferences) were considered significantly higher in quality ($d = 1.19$ [0.56–1.83]; $d = 1.10$ [0.47–1.7]). 2b. Scores on in-training examinations were comparable between lecture and FC curriculum ($d = 0.12$ [–0.4–0.6]; $d = -0.5$ [–1.0–0]; $d = -0.51$ [–1.0–0]). 3. Residents devoted more time to independent study in FC curriculum ($d = 0.85$ [0.22–1.45]).	12 (67)
Lockhart et al, ⁴⁵ 2017	Proof of concept: postevaluation	Participants evaluated the instructional format of the small private online course positively ($d = \text{NA}$).	4 (22)

While almost three-quarters used objective measures, very few articles provided validity evidence in the form of construct or content validity. Furthermore, articles investigating the relationship between relevant variables were limited. However, more than three-quarters (77%, 17 of 22) of the studies used appropriate analyses and almost 70% (15 of 22) implemented inferential statistics to analyze outcome variables such as knowledge or skills.

Discussion

This systematic review of FC application in GME yielded several important findings. While still in its early stages, research on the FC in GME has increased substantially over the past 5 years, with 17 of 22 articles being published since 2016. We discovered

use of the FC in at least 13 types of residency programs, suggesting that the teaching method can be broadly adopted across a variety of GME settings. Reasons for using the FC and specific techniques on how it was applied vary greatly across programs, suggesting that it is not a one-method-fits-all intervention. We found that the FC has been applied to entire residency programs or, on a more limited scale, within rotations or to cover specific topics. The method has also been used across institutions to prepare learners for specialized topics, or to capitalize on educator expertise.

We attribute the rise of interest in the FC in GME to medical educators' desire to find better teaching methods, recognizing that this method is suitable for residents and can be adapted to the GME setting. The self-directed learning component gives residents more

TABLE 2
Continued.

	survey only (n = 24)		
Martinelli et al, ⁴⁶ 2017	Pretest-posttest control group design with 4-month follow-up (FC = 81, control = 56)	2a. FC participant's preference for the FC significantly improved pre-post ($d = 0.95$ [0.47–1.4]). 2b. FC approach resulted in higher levels of long-term knowledge retention after 4 months ($d = 0.56$ [0.2–0.9]).	15.5 (86)
Moeller et al, ⁴⁷ 2017	Pilot study: postevaluation survey only (n = 15)	Participants evaluated the video-based lectures positively as prep for FC ($d = NA$).	7 (39)
Mokadam et al, ⁴⁸ 2016	Time series control group design (FC = 10, control = NR)	1. Participants evaluated the FC experience positively compared to lecture ($d = NA$). 2b. Change in participant scores on knowledge tests change over time at a higher rate than those of faculty ($d = NA$). 3. Reading of content material increased significantly in FC model ($d = NA$).	12 (67)
Olsen et al, ⁴⁹ 2018	Static group comparison using self-efficacy measure (FC = 17, control = 10)	2a. Participants self-efficacy increased after implementing FC with simulation boot camp: Procedures ($d = 0.77$ [0.10–1.4]); Autonomy ($d = 0.89$ [0.20–1.6]); Overall ($d = 0.84$ [0.16–1.5]).	9 (50)
Ortega et al, ⁵⁰ 2017	Postevaluation survey only (n = 25)	Participants evaluated the interactive textbook positively as prep for FC ($d = NA$).	6 (33)
Peterson et al, ⁵¹ 2017	Pilot study: Pretest-posttest control group design (FC = 10, control = 19)	1. Participants in the FC rated the experience positively ($d = NA$). 2b. Residents in the FC performed better on all 3 knowledge test topics as compared to control: Obstructive sleep apnea ($d = 0.93$ [0.09–1.8]); Acute otitis media ($d = 1.10$ [0.28–1.9]); ADHD ($d = 1.10$ [0.28–1.9]).	12 (67)
Riddell et al, ⁸ 2017	Crossover design with pre, post, and follow-up tests (FC = 37, control = 36)	2b. No difference in knowledge scores over 2 topics were observed between FC and control groups: Low back pain ($d = 0.02$ [–0.3–0.4]) and Headache ($d = 0.32$ [–0.7–1.4]).	15.5 (86)
Rose et al, ⁵² 2016	One-group pretest-posttest design with group split across 2 methods of preclass preparation (intervention = 17, control = 17)	Residents who received questions imbedded into pre-FC didactic video lectures showed significant improvement in scores compared to those who did not ($d = NA$). Residents preferred online lectures to live lectures and the inclusion of questions in online lectures ($d = NA$). Participants did not accurately report their online viewing behaviors ($d = NA$).	12 (67)
Sajedi et al, ⁵³ 2018	One-group pretest-posttest design (n = 12)	2a. Participants reported reduced anxiety and increased comfort levels about call cases after FC intervention ($d = NA$). 2b. Participants scores on knowledge tests went up after FC intervention ($d = NA$).	11 (61)
Tainter et al, ⁵⁴ 2017	One-group pretest-posttest design (n = 39)	2a. Ratings of confidence in and likelihood of using ultrasound increased significantly ($d = NA$). 2b. Post intervention scores of knowledge significantly improved on all 4 modules ($d = NA$).	12 (67)

control, allowing them to learn content at their own pace and during times outside of their clinical responsibilities. Consistent with social learning theory, the FC provides a venue for collaborative learning where learners publicly demonstrate their application of knowledge during small group discussion of patient cases and problems.^{7,8} Learner enthusiasm

for the FC method and documented improvements in knowledge and skills make the FC an attractive alternative to the traditional lecture-based education model.

Medical educators have suggested that their interest in the FC is associated with making more effective use of instructional time, providing more

TABLE 2
Continued.

Urban et al, ⁵⁵ 2016	Static group comparison (FC = 30, control = 259)	2b. Scores on in-service examination subtest covering content taught with FC increased 6.5% over same content taught by lecture during years prior ($d = 0.64 [0.3-1.0]$).	13.5 (75)
Valente et al, ⁵⁶ 2018	Static group comparison (FC = 10, control = 15)	2b. No significant differences on in-training exam were observed between FC and traditional cohorts after controlling for USMLE scores ($d = NA$).	12.5 (69)
Vasilopoulos et al, ⁵⁷ 2015	One-group pretest-posttest design with group split by levels of podcast experience. (intervention = 33, control = 24)	Residents who completed podcasts prior to guided-instruction did better than residents who did traditional didactic instruction prior to guided-instruction. ($d = 0.35 [-0.2-0.9]$). Those who had more experience with podcasts performed significantly better than those who had less experience ($d = NA$).	13 (72)
Young et al, ⁵⁸ 2014	Pilot study: postevaluation survey only (n = 35)	Participants preferred online video lectures over live lectures and the majority stated a preference for FC on a monthly basis ($d = NA$). 1. Participants offered more advantages for using the FC than they did disadvantages ($d = NA$).	6 (33)

^a Cohen's d effect sizes are generally interpreted as < 0.1 , no effect; $0.1-0.4$, small effect; $0.5-0.7$, intermediate effect; and > 0.7 , large effect.

Abbreviations: MERSQI, Medical Education Research Study Quality Instrument; NA, not available; FC, flipped classroom; QIKAT, Quality Improvement Knowledge Application Tool; TBL, team-based learning; NR, not reported; QI, quality improvement; ADHD, Attention-Deficit/Hyperactivity Disorder.

Note: Dark shading indicates studies that focus predominantly on preclass methods rather than the FC method.

structure for self-directed learning outside of class, or motivating residents to spend more time outside of clinical responsibilities engaged in study. The FC requires learners to complete self-directed learning activities to participate during in-class sessions. This was a concern among graduate medical educators because residents have clinical responsibilities that compete for their time. We believe many studies created more interesting methods for delivering preclassroom content because of this concern.

F. Chen and colleagues' earlier systematic review of the FC approach in medical education found that the "majority of literature has been carried out in UME."³ Additionally, their review suggested that research up to that point "lacked strong evidence for the effectiveness of FCs."³ A subsequent comprehensive meta-analysis by K.S. Chen et al covered a relatively small number of studies involving medical education and only one that involved residency education.³⁰ This group of authors tentatively suggested an advantage of the FC over lecture-based methods; however, they expressed concern about this interpretation due to extreme diversity in methods. Our systematic review focused solely on the use of the FC in ACGME-accredited residency programs. Like K.S. Chen, we confronted diversity in research methods, but also diversity in how the FC was applied and the content covered. In contrast with F. Chen, and comparable to K.S. Chen, we found that learners generally find the FC approach acceptable and evidence supports that the FC is as good as a traditional didactic approach (introducing topics

through lecture or during a face-to-face meeting followed by readings and study of educational materials).

This systematic review of the FC literature has some limitations. While we employed what we think is a comprehensive search strategy with the help of an experienced medical librarian, it is possible that we did not include more esoteric terms that refer to methods associated with FCs. For example, we did not use the term "problem-based learning," which is a method that typically describes a comprehensive approach to education, but could be associated with the FC format on a more limited basis. In addition, we made the decision to exclude non-ACGME residency programs and US fellowship programs. This decision was based on the variability of education structures found outside the United States and variability in size and purpose of fellowships. These decisions may have restricted the generalizability of our findings.

The assessment of higher-level outcomes, such as changes in learner behaviors or patient outcomes, remains a challenge in medical education.³⁶ These outcomes are difficult to assess because they rarely can be attributed to a single educational intervention. Despite these challenges, we were heartened to find an increasing number of studies that generated Kirkpatrick levels 2a, 2b, and 3 outcomes, many studies that employed objective measures, and an increasing number of studies that employed more rigorous research designs. Future research on the FC in GME should focus on higher-level outcomes such as

changes in behaviors, clinical practice, and patient outcomes.³³

Conclusion

The FC pedagogical approach in GME has been implemented in a variety of ways and studied with a variety of methods, which has yielded variable results. Using MERSQI scores, studies evaluating the efficacy of the FC were somewhat less rigorous than typical medical education research studies; however, if pilot and proof-of-concept studies are eliminated, the average MERSQI score was comparable to that of other medical education research studies. The studies that evaluated resident satisfaction or efficacy concluded that residents held generally positive opinions about the FC. For studies that evaluated learning outcomes, the results were mixed: slightly more than half of the studies using a control group for comparison found positive learning results. Future studies of FC in GME should include higher-level outcomes (changes in knowledge, behaviors, or patient outcomes) and assessment of preclassroom assignment completion.

References

1. Mehta NB, Hull AL, Young JB, Stoller JK. Just imagine: new paradigms for medical education. *Acad Med*. 2013;88(10):1418–1423. doi:10.1097/ACM.0b013e3182a36a07.
2. Straumsheim C. Still in favor of the flip. October 30, 2013. Inside Higher Education. <https://www.insidehighered.com/news/2013/10/30/despite-new-studies-flipping-classroom-still-enjoys-widespread-support>. Accessed December 21, 2018.
3. Chen F, Lui AM, Martinelli SM. A systematic review of the effectiveness of flipped classrooms in medical education. *Med Educ*. 2017;51(6):585–597. doi:10.1111/medu.13272.
4. King A, Boysen-Osborn M, Cooney R, Mitzman J, Misra A, Williams J, et al. Curated collection for educators: five key papers about the flipped classroom methodology. *Cureus*. 2017;9(10):e1801. doi:10.7759/cureus.1801.
5. Cooper AZ, Hsieh G, Kiss JE, et al. Flipping out: does the flipped classroom learning model work for GME? *J Grad Med Educ*. 2017;9(3):392–393. doi:10.4300/JGME-D-16-00827.1.
6. Vygotsky LS. *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press; 1978.
7. Hawks SJ. The flipped classroom: now or never? *AANA J*. 2014;82(4):264–269.
8. Riddell J, Jhun P, Fung CC, Comes J, Sawtelle S, Tabatabai R, et al. Does the flipped classroom improve learning in graduate medical education? *J Grad Med Educ*. 2017;9(4):491–496. doi: 10.4300/JGME-D-16-00817.1.
9. Sherbino J, Chan T, Schiff K. The reverse classroom: lectures on your own and homework with faculty. *CJEM*. 2013;15(3):178–180.
10. Bloom BS, Engelhart MD, Furst EJ, et al. The classification of educational goals. In: Bloom BS, ed. *Taxonomy of Educational Objectives. Handbook I: Cognitive Domain*. New York, NY: David McKay; 1956.
11. Prober CG, Khan S. Medical education reimaged: a call to action. *Acad Med*. 2013;88(10):1407–1410. doi:10.1097/ACM.0b013e3182a368bd.
12. McLaughlin JE, Roth MT, Glatt DM, Gharkholonarehe N, Davidson CA, Griffin LM, et al. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. *Acad Med*. 2014;89(2):236–243. doi:10.1097/ACM.0000000000000086.
13. Leung JY, Kumta SM, Jin Y, Yung AL. Short review of the flipped classroom approach. *Med Educ*. 2014;48(11):1127. doi:10.1111/medu.12576.
14. Nematollahi S, St John PA, Adamas-Rappaport WJ. Lessons learned with a flipped classroom. *Med Educ*. 2015;49(11):1143. doi:10.1111/medu.12845.
15. Morgan H, McLean K, Chapman C, Fitzgerald J, Yousuf A, Hammoud M. The flipped classroom for medical students. *Clin Teach*. 2015;12(3):155–160. doi:10.1111/tct.12328.
16. Vincent DS. Out of the wilderness: flipping the classroom to advance scholarship in an internal medicine residency program. *Hawaii J Med Public Health*. 2014;73(11 suppl 2):2–3.
17. Ramar K, Hale CW, Dankbar EC. Innovative model of delivering quality improvement education for trainees—a pilot project. *Med Educ Online*. 2015;20:28764. doi:10.3402/meo.v20.28764.
18. Sadosty AT, Goyal DG, Hern HG Jr, Kilian BJ, Beeson MS. Alternatives to the conference status quo: summary recommendations from the 2008 CORD Academic Assembly Conference Alternatives workgroup. *Acad Emerg Med*. 2009;16(suppl 2):25–31. doi:10.1111/j.1553-2712.2009.00588.x.
19. Wittich CM, Agrawal A, Wang AT, Halvorsen AJ, Mandrekar JN, Chaudhry S, et al. Flipped classrooms in graduate medical education: a national survey of residency program directors. *Acad Med*. 2018;93(3):471–477. doi:10.1097/ACM.0000000000001776.
20. Bergmann J. *Flip Your Classroom: Reach Every Student in Every Class Every Day*. Arlington, VA: International Society for Technology in Education; 2012.

21. Muzyk AJ, Fuller S, Jiroutek MR, Grochowski CO, Butler AC, Byron May D. Implementation of a flipped classroom model to teach psychopharmacotherapy to third-year doctor of pharmacy (PharmD) students. *Pharm Educ*. 2015;15(1):44–53.
22. Critz CM, Knight D. Using the flipped classroom in graduate nursing education. *Nurse Educ*. 2013;38(5):210–213. doi:10.1097/NNE.0b013e3182a0e56a.
23. Ferreri SP, O'Connor SK. Redesign of a large lecture course into a small-group learning course. *Am J Pharm Educ*. 2013;77(1):13. doi:10.5688/ajpe77113.
24. McGowan BS, Balmer JT, Chappell K. Flipping the classroom: a data-driven model for nursing education. *J Contin Educ Nurs*. 2014;45(11):477–478. doi:10.3928/00220124-20141027-11.
25. Simpson V, Richards E. Flipping the classroom to teach population health: increasing the relevance. *Nurse Educ Pract*. 2015;15(3):162–167. doi:10.1016/j.nepr.2014.12.001.
26. Pierce R, Fox J. Vodcasts and active-learning exercises in a “flipped classroom” model of a renal pharmacotherapy module. *Am J Pharm Educ*. 2012;76(10):196. doi:10.5688/ajpe7610196.
27. Singleton JA, Nissen LM. Teaching pharmacy students how to manage effectively in a highly competitive environment. *Pharm Educ*. 2014;14(1):21–25.
28. Wong TH, Ip EJ, Lopes I, Rajagopalan V. Pharmacy students' performance and perceptions in a flipped teaching pilot on cardiac arrhythmias. *Am J Pharm Educ*. 2014;78(10):185. doi:10.5688/ajpe7810185.
29. Moffet J, Mill AC. Evaluation of the flipped classroom approach in a veterinary professional skills course. *Adv Med Educ Pract*. 2014;5:415–425. doi:10.2147/AMEP.S70160.
30. Chen KS, Monrouxe L, Lu YH, Jenq CC, Chang YJ, Change YC, et al. Academic outcomes of flipped classroom learning: a meta-analysis. *Med Educ*. 2018 Jun 25. doi:10.1111/medu.13616. Epub ahead of print.
31. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269.
32. Gay LR. *Educational Research: Competencies for Analysis and Application*. 2nd ed. Columbus, OH: Charles E. Merrill Publishing Co; 1981.
33. Boet S, Sharma S, Goldman J, Reeves S. Review article: medical education research: an overview of methods. *Can J Anesth* 2012;59(2):159–170. doi:10.1007/s12630-011-9635-y.
34. Lenhard W, Lenhard A. Calculation of effect sizes. Dettelbach, Germany: Psychometrica; 2016. https://www.psychometrica.de/effect_size.html. doi:10.13140/RG.2.1.3478.4245.
35. Wilson DB. Practical meta-analysis effect size calculator. <http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>. Accessed December 21, 2018.
36. Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. *Acad Med*. 2015;90(8):1067–1076. doi:10.1097/ACM.0000000000000786.
37. Reed DA, Cook CA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002–9.
38. Barrie M, Amick C, Mitzman J, Way DP, King AM. Flipping the emergency medicine intern orientation classroom. *Western J Emerg Med*. 2018;19(1):145–147.
39. Bonnes SL, Ratelle JT, Halvorsen AJ, Carter KJ, Hafdahl LT, Wang AT, et al. Flipping the quality improvement classroom in residency education. *Acad Med*. 2017;92(1):101–107. doi:10.1097/ACM.0000000000001412.
40. Chokshi BD, Schumacher HK, Reese K, Bhansali P, Kern JR, Simmens SJ, et al. A “resident-as-teacher” curriculum using a flipped classroom approach: can a model designed for efficiency also be effective? *Acad Med*. 2017;92(4):511–514. doi:10.1097/ACM.0000000000001534.
41. Girgis F, Miller JP. Implementation of a “flipped classroom” for neurosurgery resident education. *Can J Neurol Sci*. 2018;45(1):76–82. doi:10.1017/cjn.2017.234.
42. Haspel RL, Ali AM, Huang GC. Using a team-based learning approach at national meetings to teach residents genomic pathology. *J Grad Med Educ*. 2016;8(1):80–84. doi:10.4300/JGME-D-15-00221.1.
43. Keefer P, Orringer K, Vredevel J, Warriar K, Burrows H. Developing a quality improvement and patient safety toolbox: the curriculum. *MedEdPORTAL*. 2016;12:10385. https://doi.org/10.15766/mep_2374-8265.10385.
44. King AM, Mayer C, Barrie M, Greenberger S, Way DP. Replacing lectures with small groups: the impact of flipping the residency conference day. *West J Emerg Med*. 2018;19(1):11–17. doi: 10.5811/westjem.2017.10.35235.
45. Lockhart BJ, Capurso NA, Chase I, Arbuckle MR, Travis MJ, Eisen J, et al. The use of a small private online course to allow educators to share teaching resources across diverse sites: the future of psychiatric case conferences? *Acad Psychiatry*. 2017;41(1):81–85. doi:10.1007/s40596-015-0460-4.
46. Martinelli SM, Chen F, DiLorenzo AN, Mayer DC, Fairbanks S, Moran K, et al. Results of a flipped

- classroom teaching approach in anesthesiology residents. *J Grad Med Educ*. 2017;9(4):485–490. doi:10.4300/JGME-D-17-00128.1.
47. Moeller JJ, Farooque P, Leydon G, Dominguez M, Schwartz ML, Sadler RM. A video-based introductory EEG curriculum for neurology residents and other EEG learners. *MedEdPORTAL*. 2017;13:10570.
 48. Mokadam NA, Dardas TF, Hermesen JL, Pal JD, Mulligan MS, Jacobs LM, et al. Flipping the classroom: case-based learning, accountability, assessment, and feedback leads to a favorable change in culture. *J Thorac Cardiovasc*. 2016;153(4):987–996.e1. doi:10.1016/j.jtcvs.2016.10.101.
 49. Olsen KR, Bannister L, Deshmukh A, Hall DJ, Mira J, Patel R, et al. Simulation-based learning improves anesthesiology resident self-efficacy in critical skills: a flipped classroom approach. *Med Sci Educ*. 2018;28(1):65–69. doi:10.1007/s40670-017-0497-y.
 50. Ortega R, Akhtar-Khavari V, Barash P, Sharar S, Stock MC. An innovative textbook: design and implementation. *Clin Teach*. 2017;14(6):407–411. doi:10.1111/tct.12587.
 51. Peterson J, Loudon DT, Gribben V, Blankenburg R. Teaching residents clinical practice guidelines using a flipped classroom model. *MedEdPORTAL*. 2017;13:10549.
 52. Rose E, Claudius I, Tabatabai R, Kearl L, Behar S, Jhun P. The flipped classroom in emergency medicine using online videos with interpolated questions. *J Emerg Med*. 2016;51(3):284–291.e1. doi:10.1016/j.jemermed.2016.05.033.
 53. Sajedi P, Salamon N, Hostetter J, Karnezis S, Vijayasarathi A. Reshaping radiology precall preparation: integrating a cloud-based PACS viewer into a flipped classroom model. *Curr Probl Diagn Radiol*. 2018 Jul 29. pii:S0363-0188(18)30153-1. doi:10.1067/j.cpradiol.2018.07.014. Epub ahead of print.
 54. Tainter CR, Wong NL, Cudemus-Deseda GA, Bittner EA. The “flipped classroom” model for teaching in the intensive care unit: rationale, practical considerations, and an example of successful implementation. *J Intensive Care Med*. 2017;32(3):187–196. doi:10.1177/0885066616632156.
 55. Urban RR, Swensen RE, Schulkin J, Schiff MA. Implementing the “flipped classroom” on a gynecologic oncology service. *J Reprod Med*. 2016;61(9–10):405–410.
 56. Valente A, Gala R, Rueb B. Analysis of OBGYN resident in-training exam scores after implementation of a flipped classroom curriculum. *MedEdPublish*. 2018;7(2):24. doi:https://doi.org/10.15694/mep.2018.0000092.1.
 57. Vasilopoulos T, Chau DF, Bensalem-Owen M, Cibula JE, Fahy BG. Prior podcast experience moderates improvement in electroencephalography evaluation after educational podcast module. *Anesth Analg*. 2015;121(3):791–797. doi:10.1213/ANE.0000000000000681.
 58. Young TP, Bailey CJ, Guptill M, Thorp AW, Thomas TL. The flipped classroom: a modality for mixed asynchronous and synchronous learning in a residency program. *West J Emerg Med*. 2014;15(7):938–944. doi:10.5811/westjem.2014.10.23515.



Andrew M. King, MD, is Associate Professor, Department of Emergency Medicine, The Ohio State University; **Michael Gottlieb, MD**, is Assistant Professor, Department of Emergency Medicine, Rush University Medical Center; **Jennifer Mitzman, MD**, is Assistant Professor, Department of Emergency Medicine, The Ohio State University; **Tina Dulani, MD**, is Assistant Professor, Department of Emergency Medicine, Hofstra North Shore-LIJ; **Stephanie J. Schulte, MLIS**, is Associate Professor, Health Sciences Library in Prior Hall, The Ohio State University; and **David P. Way, MEd**, is Medical Education Research Specialist, Department of Emergency Medicine, The Ohio State University.

Funding: The authors report no external funding for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Andrew M. King, MD, The Ohio State University, Department of Emergency Medicine, 760 Prior Hall, 376 West 10th Avenue, Columbus, OH 43210, 614.293.3551, andrew.king3@osumc.edu

Received May 7, 2018; revisions received October 11, 2018, and November 17, 2018; accepted December 11, 2018.