

# The Effect of a Flipped Classroom Learning Model Versus Traditional Lecture Model on Resident's Knowledge Acquisition for Atraumatic Back Pain in the Emergency Department

Journal of Medical Education and Curricular Development  
Volume 10: 1–5  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/23821205231193283



Elizebeth Dubey<sup>1</sup> , James H. Paxton<sup>1</sup>, Laura Smylie<sup>1</sup>, Robert D. Welch<sup>1</sup> and Anne Messman<sup>1</sup>

<sup>1</sup>Department of Emergency Medicine, Wayne State University School of Medicine, Detroit, MI, USA.

## ABSTRACT

**OBJECTIVE:** The effectiveness of lecture-based (LB) formats for residency education has recently been challenged as the gold standard. Studies suggest a flipped classroom (FC) lecture improves resident satisfaction, but evidence that showing improved knowledge acquisition is lacking. To determine whether the flipped classroom model improves knowledge acquisition compared to traditional LB model.

**METHODS:** Emergency medicine resident physicians at 2 academic programs were included in December 2019; at Sinai-Grace Hospital, a traditional lecture was the teaching method and at Detroit Receiving Hospital, FC was utilized. Residents completed prelecture and postlecture content tests. The primary outcome was change in test results (pretest to post-test). A noninferiority design comparing the changes between intervention and control groups was utilized (1-sided t-test, noninferiority margin of  $-0.5$ ; 1-sided  $\alpha = 0.05$ ).

**RESULTS:** Results were available for 31 residents (17 controls and 14 interventions) out of 83 enrolled. There were 14 postgraduate year 1 (PGY-1), 9 PGY-2, and 8 PGY-3 residents. The mean difference in score was  $+0.71$  (SD 1.38) and  $+0.77$  (SD 1.48) for the FC and LB groups, respectively. This resulted in a mean difference between groups of  $-0.05$  (lower bound of the upper 95% confidence interval  $-0.93$  and therefore crossing the noninferiority margin of  $-0.05$ ;  $P = .20$ ).

**CONCLUSIONS:** This study of resident education at 2 training programs was unable to demonstrate noninferiority of an FC format compared to standard lecture. Surprisingly, there was little improvement in test results after both teaching formats. Larger studies are needed to power results.

**KEYWORDS:** Flip classroom, medical education, low back pain

**RECEIVED:** May 13, 2022. **ACCEPTED:** July 3, 2023

**TYPE:** Original Research Article

**FUNDING:** The authors received no financial support for the research, authorship, and/or publication of this article.

**DECLARATION OF CONFLICTING INTERESTS:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Elizebeth Dubey, Department of Emergency Medicine, Wayne State University School of Medicine, 4201 St. Antoine St, STE 6F UHC, Detroit, MI 48201, USA. Email: edube@med.wayne.edu

## Introduction

Grand Round conferences in emergency medicine residency training are an integral part of residency education as they contribute to teaching residents knowledge, competency, and the practice of emergency medicine. It is essential to develop and employ curriculum to maximize resident learning while balancing time invested by both learners and faculty. Lecture-based format has traditionally been the primary teaching method for graduate medical education.<sup>1</sup> In the lecture-based (LB) format, the responsibility for teaching material is placed on the teacher while learners assimilate information, and thus knowledge acquisition is theorized to occur during the educational session.

Recently, the effectiveness of the LB format as an ideal teaching method has been questioned due to lack of learner engagement.<sup>2</sup> There has been a recent shift to replace lectures with other methods that may promote more active learning.<sup>3–5</sup> The flipped classroom (FC) teaching method has been used in graduate medical education for some years, and has recently become an increasingly popular teaching method in medical education.<sup>6,7</sup> In a flipped classroom format, content attainment

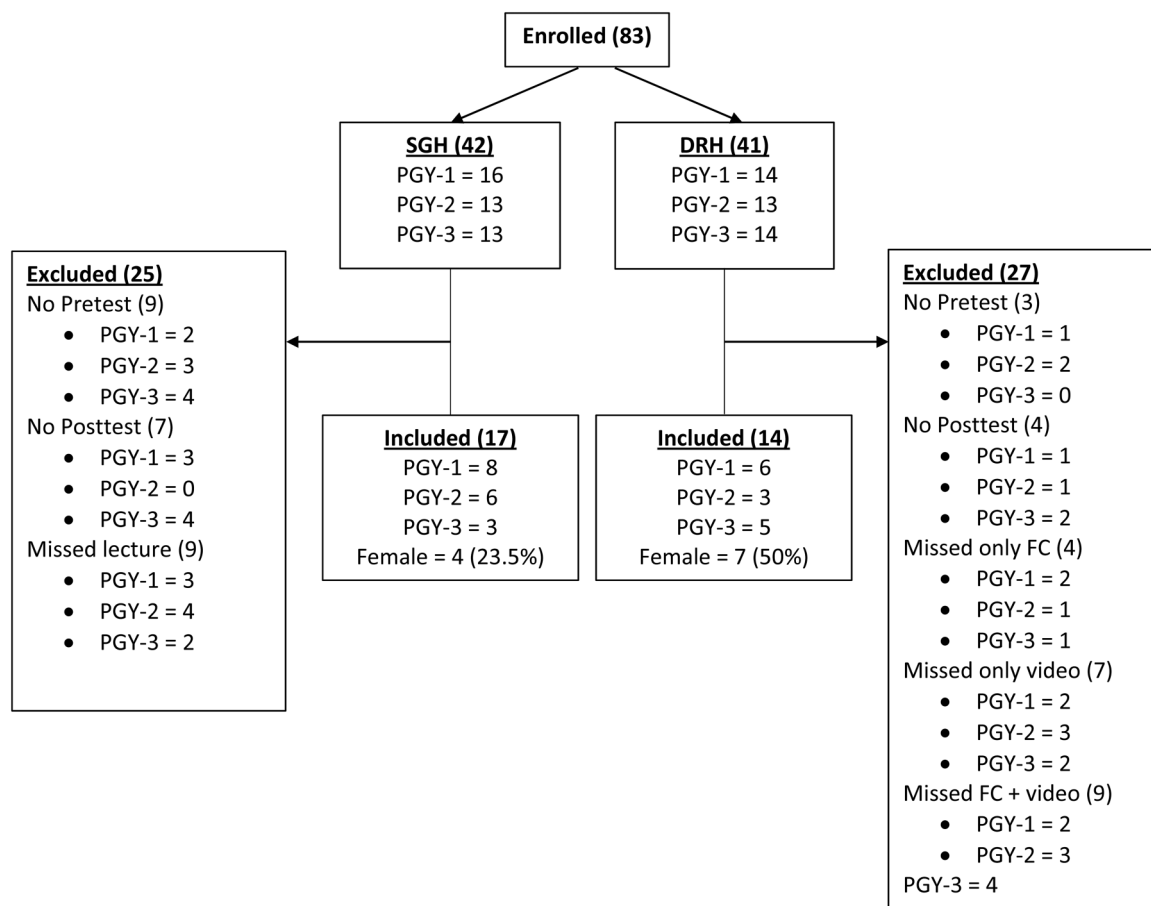
precedes the classroom via an online format, and class time is spent engaging students in learning activities that allow for collaboration, discussion and interaction among students.<sup>8</sup> While there are many avenues of employment, the learner typically processes reading, videos, or other learning materials prior to class and then engages in discussions and problem solving to advance understanding of the content during class time. Asynchronous learning via video lectures and practice problems followed by active, group-based problem-solving activities is an important component of the FC method and what differentiates it from other similar instructional activities.<sup>9</sup> The learning is believed to be more student-centered, with substantial knowledge acquisition occurring prior to the classroom session. In the classroom, more in-depth application-based learning, when facilitated by the teacher, is thought to be achievable.<sup>10</sup>

While the FC model has gained popularity, there is little evidence that it affects higher-level learning outcomes based on Kirkpatrick's model.<sup>9</sup> Recent reviews have shown that although students "like" FC, there are mixed results regarding improvement in higher level outcomes such as examination scores.<sup>9</sup> Our



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).



**Figure 1.** Flow diagram. Abbreviations: DRH, Detroit Receiving Hospital; FC, flipped classroom; Lecture, in-conference lecture; PGY, postgraduate year; SGH, Sinai Grace Hospital; Video, recorded video.

pilot study creates a framework to evaluate differences in knowledge acquisition between emergency medicine (EM) residents who were taught via traditional classroom model versus FC model. We hypothesized that an FC curriculum versus a traditional lecture curriculum will not be associated with a difference in knowledge acquisition among a categorical EM resident population.

## Methods

The population of interest included resident physicians in training at 2 distinct emergency medicine residencies, both affiliated with Wayne State University, Detroit, Michigan for the academic year 2019 to 2020. *Informed consent from study subjects were waived by the Wayne State University Institutional Review Board (see above for IRB).* Both residency programs, *Sinai-Grace Hospital (SGH) and Detroit Receiving Hospital (DRH), are 3-year categorical programs.* Resident physicians from all 3 years of residency training were included in the study population. All residents were eligible to participate in the study but were excluded if they did not complete the pretest, the intervention, or the post-test. A total of 83 categorical residents were eligible for participation across the 2 sites.

During EM weekly didactics, residents were educated on the topic of atraumatic back pain. This topic was chosen from a list of 10 topics listed as part of the American College of Emergency Physicians “Choosing Wisely” campaign.<sup>11</sup> A faculty member not affiliated with the study, and without knowledge of the purpose of the study, prepared and presented both the FC style lecture and the traditional lecture in December 2019. Residents at SGH received the traditional lecture format and residents at DRH received the FC style lecture. Both groups were given material to review beforehand via their online residency education website with the expectation that residents review the material prior to the conference. Examples of what were used were textbook reading such as Harwood and Nuss, Chapter 148: Low Back Pain and Rosen’s Emergency Medicine, Chapter 43: Musculoskeletal Back pain; online FOAMED like FOAMcast: Back pain and spinal epidural abscess; articles like EB Medicine’s “An Evidence-Based Approach To The Evaluation And Treatment Of Low Back Pain In The Emergency Department.” The DRH residents had access to the same review material as the SGH residents on their residency website, with the addition of the video recording of the

**Table 1.** Demographic data for participants.

Site/level of training	Phase of study	
	Enrolled <sup>a</sup> n (%)	Completed study <sup>b</sup> n (%)
SGH	42 (50.6)	17 (40.5)
PGY-1	16 (38.1)	8 (50.0)
PGY-2	13 (31.0)	6 (46.2)
PGY-3	13 (31.0)	3 (23.1)
Female	8 (19.1)	4 (50.0)
DRH	41 (49.4)	14 (34.2)
PGY-1	14 (34.2)	6 (42.9)
PGY-2	13 (31.7)	3 (23.1)
PGY-3	14 (34.2)	5 (35.7)
Female	17 (41.5)	7 (41.2)
Total	83 (100.0)	31 (37.4)
PGY-1	30 (36.2)	14 (46.7)
PGY-2	26 (31.3)	9 (34.6)
PGY-3	27 (32.5)	8 (29.7)
Female	25 (30.1)	11 (44.0)

<sup>a</sup>Enrolled participants are those who completed either the pretest or post-test.

<sup>b</sup>Completed study were those participants who completed the module, as well as both the pretest and post-test.

Abbreviations: DRH, Detroit Receiving Hospital; PGY, postgraduate year; SGH, Sinai Grace Hospital.

traditional back pain lecture that was given to the SGH residents the week prior. Both residencies were exposed to the topic material the same number of times (before and during conference) so that number of content exposures between residency programs would not be a variable that affects the study endpoints. The only difference was the addition of the traditional back pain lecture video for DRH as part of the pre-review. In the DRH conference, the material was discussed in depth via instructor-facilitated concept application activities. The structure of the FC session was at the discretion of the faculty lecturer and could have been case-based learning, algorithm building, pairing games, etc. The lecturer chose to employ small group case-based learning stations for this topic.

The demographic information collected during the written test included: (1) postgraduate year (PGY) of training, (2) gender, and (3) the resident's clinical site (SGH or DRH). Providing this information was optional and residents could complete the test without providing this demographic information. Provision of this information was required to ensure the integrity of data, however, was removed prior to data analysis. The data was matched and de-identified by a third-party

faculty member prior to distribution to the remainder of the study team. The data was stored on a hard drive in the private office of the third-party faculty member.

Testing instruments were distributed to residents via e-mail link to SurveyMonkey®. The test contained 5 questions requiring short write-in answers. These tests were created by the study authors and content validated by experts with significant experience in medical education research to ensure clarity (Supplemental Appendix 1). Each test item was worth a designated number of points based on the complexity of the question, with a maximum total test score of 10 points. Testing was administered 2 weeks before (pretest) and 2 weeks after (post-test) the educational intervention in December 2019. After completion, the responses were collected by a research team member who had no part in scoring. This research team member de-identified the responses and assigned a unique identifier. Unique identifiers from pre-intervention and post-intervention datasets were compared to verify the same participants in each group. Comparison was made across the test results by the research team member.

### Statistical Analysis

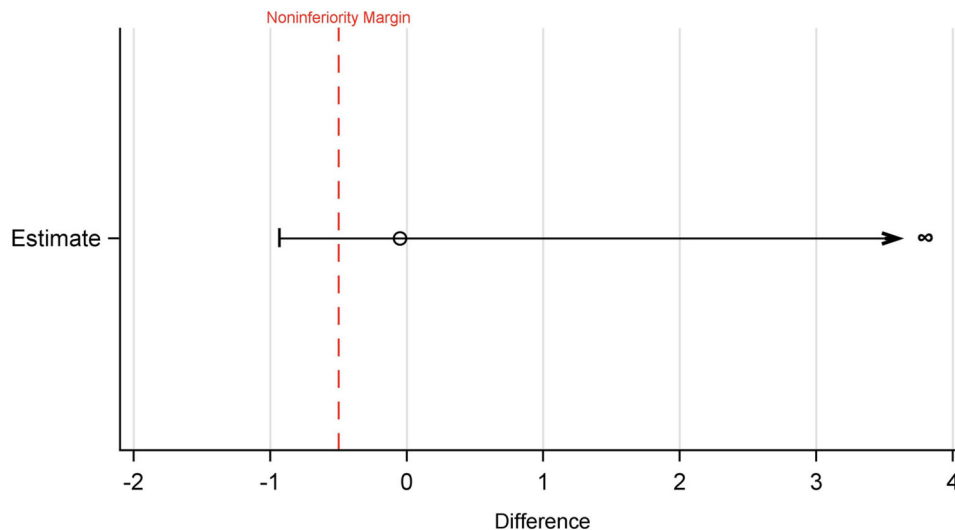
De-identified responses were then scored by 2 independent reviewers. Data analysis was performed using standard techniques. Continuous variables were reported as mean with standard deviation or median with interquartile range where appropriate. Categorical variables were reported as percentages. Cohen's kappa ( $\kappa$ ) coefficient for inter-rater reliability was used to compare the 2 study authors scoring the written pretest and post-test. For the difference in the change in test scores (post-test minus pretest), a 1-sided 2-sample t-test was utilized. We selected a noninferiority margin of  $-0.5$  a priori and considered a 1-side  $P$  value  $< .05$  as significant. Since this was an early observational study designed to provide data for future work, a sample size estimation was not used. This study was approved by our institutional review board.

### Results

Eighty-three residents participated in the study, including 42 SGH residents and 41 DRH residents. After excluding those residents who failed to complete one of the tests or attend the intervention, a total of 31 (37.3%) respondents were included in the final analysis as seen in the CONSORT diagram in Figure 1. Demographic data for participants is provided in Table 1.

The observed Cohen's kappa for all responses was 0.78 (95% confidence interval [CI] 0.70-0.88), with an observed agreement of 91.9% for all study questions.

For the main outcome of change in test scores, the mean difference in score (ie, [post-test score] – [pretest score]) for the FC group was +0.71 (SD 1.38), and +0.77 (SD 1.48) for the LB group, representing a mean increase in correct answers of  $< 1$  in both study groups. Nine (64.3%) of the FC participants



**Figure 2.** Mean differences. Mean of Difference in Differences (FC – Control)\* (with Upper 95% CI). Mean change in test scores demonstrating noninferiority of the FC method. Abbreviation: FC, flipped classroom.

had a positive score increase after the intervention. Eleven (64.7%) of the LB participants had a positive score increase.

The difference in the mean change in test scores between groups was 0.05 (SD 1.44) therefore crossing the noninferiority margin of  $-0.5$  ( $P = .20$ ). Figure 2 shows the upper 95% CI crossing the specified noninferiority boundary.

## Discussion

This study was designed to determine whether an FC model is noninferior to a LB model of educational content delivery as it relates to resident knowledge acquisition in a single educational intervention. We found no significant difference in pre-intervention and postintervention test scores between the 2 study groups. Using a noninferiority margin of  $-0.5$ , we were unable to demonstrate noninferiority.

Our research adds to increasing evidence that while FC is associated with increased learner satisfaction, results are mixed in demonstrating its effectiveness in promoting knowledge acquisition above a traditional LB format. A recent review by Chen et al examined many relevant studies to explore their benefits based on Kirkpatrick's model. The effectiveness of the FC was classified by Kirkpatrick's model ranging from outcomes that change learner perception (level 1) to those that affect patient outcomes (level 4). Findings from this review suggest that the FC is a favorable teaching approach when the intent is to increase learners' motivation, task value, and engagement (level 1), and though results were mixed with regard to knowledge, the FC was shown to be at least as effective as traditional education.<sup>9</sup>

One possibility for the lack of a difference seen in our study is that both study groups were taught by the same well-prepared instructor who engaged learners in active learning. A review done by Kraut et al assessed the benefits and drawbacks of an

FC approach and concluded that while the FC model may be better for learning higher cognition tasks, learner satisfaction depended largely on teacher preparatory work.<sup>12</sup>

Our study has several limitations. While the test questionnaires were content validated by study authors, they were not pilot tested prior to administration. Only 1 post-test was performed 2 weeks following the intervention, and we did not analyze knowledge retention beyond this point. It's possible that difference between the study groups would have been measurable months after the intervention. Also, due to the constraints of those included in the data analysis (the flip arm must have watched the video, be present for lecture/flip, and have taken both the pretest and post-test), our overall participant numbers were low. Both of our quizzes were anonymous and voluntary, which led to difficulty recruiting all potential study participants. Because of this, our study was not powered for subgroup analysis and it's unclear if trainees of different PGY levels or other demographics performed differently from one another. A sample size calculation was not used, and it is possible that a larger sample would alter our results.

Our study only compared 2 separate residency programs in a Midwest urban training environment. A similarly structured study involving a greater number and variety of emergency medicine residency programs would likely yield more robust results. In addition, we plan to use the structural framework of this study to examine educational outcomes of resident physicians using other educational topics. Lastly, this study looked at knowledge acquisition which is level 2 on Kirkpatrick's model.<sup>13</sup> While studies assessing learner satisfaction (level 1) and knowledge (level 2) have been the bulk of research on FC thus far, the next logical step is to go beyond lower-level outcomes and focus on learner behavior, or in the case of residents, clinical behavior.

## Conclusion

In conclusion, our study sought to determine whether the FC model improves knowledge acquisition for learners when compared to a traditional lecture and showed that FC was noninferior to traditional lecture.

## Ethics Approval

This study was approved by Wayne State University's IRB on 6/24/2019, IRB# 051819B3X.

## ORCID iD

Elizebeth Dubey  <https://orcid.org/0000-0001-6754-1029>

## Supplemental Material

Supplemental material for this article is available online.

## REFERENCES

1. Cooper AZ, Richards JB. Lectures for adult learners: breaking old habits in graduate medical education. *Am J Med.* 2017;130(3):376-381.
2. Hurst JW. The overlecturing and underteaching of clinical medicine. *Arch Intern Med.* 2004;164(15):1605-1608.
3. Hatem CJ, Searle NS, Gunderman R, et al. The educational attributes and responsibilities of effective medical educators. *Acad Med.* 2011;86(4):474-480.
4. Prober CG, Heath C. Lecture halls without lectures—a proposal for medical education. *N Engl J Med.* 2012;366(18):1657-1659.
5. Prober CG, Khan S. Medical education reimaged: a call to action. *Acad Med.* 2013;88(10):1407-1410.
6. Khanova J, Roth MT, Rodgers JE, McLaughlin JE. Student experiences across multiple flipped courses in a single curriculum. *Med Educ.* 2015;49(10):1038-1048.
7. Moraros J, Islam A, Yu S, Banow R, Schindelka B. Flipping for success: evaluating the effectiveness of a novel teaching approach in a graduate level setting. *BMC Med Educ.* 2015;15:27.
8. Heng Ngee M. Teaching tip: the flipped classroom. *J Inform Syst Educ.* 2014;25(1):7-11.
9. 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.
10. Moffett J. Twelve tips for “flipping” the classroom. *Med Teach.* 2015;37(4):331-336.
11. Directors ABo. ACEP//Homepage. <https://www.acep.org/administration/quality/acep-announces-list-of-tests-as-part-of-choosing-wisely-campaign>. Published 2013. Accessed June 26, 2023.
12. Kraut AS, Omon R, Caretta-Weyer H, et al. The flipped classroom: a critical appraisal. *West J Emerg Med.* 2019;20(3):527-536.
13. Kirkpatrick DL, Kirkpatrick JD, eBook Academic Collection – North America, Skillsoft Books BusinessPro Collection (Skillport Platform). Evaluating training programs the four levels. In: 3rd ed. San Francisco, CA: Berrett-Koehler; 2006: <http://proxy.lib.wayne.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=e000xna&AN=260710>. <https://proxy.lib.wayne.edu/sso/skillport?context=11864>.