

ARTICLE

Rigor and reproducibility training for first year medical students in research pathways

Kate L. Lapane | Catherine E. Dube

Division of Epidemiology, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, Massachusetts, USA

Correspondence

Kate Lapane, Division of Epidemiology, Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA 01655, USA.
Email: kate.lapane@umassmed.edu

Funding information

Funding for the curriculum development and evaluation was provided by an NIH administrative supplement to the UMMS MSTP Program (T32 GM107000-07S1) and Clinical and Translational Science Award (5UL1TR001453).

Abstract

In the Spring of 2020, we launched a rigor and reproducibility curriculum for medical students in research training programs. This required class consisted of eight, 2-h sessions, which transitioned to remote learning in response to the coronavirus disease 2019 (COVID-19) epidemic. The class was graded as pass/fail. Flipped classroom techniques, with multiple hands-on exercises, were developed for first-year medical students (MD/PhD [$n = 9$], Clinical and Translational Research Pathway (CTRP) students [$n = 9$]). Four focus groups ($n = 13$ students) and individual interviews with the two instructors were conducted in May 2020. From individual interviews with instructors and focus groups with medical students, the course and its components were favorably reviewed. Students thought the course was novel, important, relevant, and practical—and teaching strategies were effective (e.g., short lectures, interactive small group exercises, and projects). Most students expressed concerns about lack of time for course preparation. Sharper focus and streamlining of preparation work may be required. Pre- and post-student self-assessments of rigor and reproducibility competencies showed average post-scores ranging from high/moderate to strong understanding ($n = 11$). We conclude that rigor and reproducibility can be taught to first-year medical students in research pathways programs in a highly interactive and remote format.

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

The rigor and reproducibility crisis calls for robust training of scientists in best practices for enhancing the research rigor.

WHAT QUESTION DID THIS STUDY ADDRESS?

We evaluated a curriculum to develop physician-scientists skilled at documenting research workflow from idea generation to publication with reproducibility in mind.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

Highly interactive exercises, coupled with a hands-on replication group project provide a pathway for students to gain competencies important to the improvement of rigor and reproducibility in scientific research. Rigor and reproducibility can be taught in a highly interactive format and using a remote format.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Clinical and Translational Science* published by Wiley Periodicals LLC on behalf of the American Society for Clinical Pharmacology and Therapeutics.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

Formal training is needed to raise awareness of the reproducibility crisis and improve the rigor of research conducted. If techniques taught are used, the transparency and reproducibility of clinical and translational science will be improved.

INTRODUCTION

Reproducibility is the ability “to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator.”^{1,2} When results of a prior study are duplicated using the same procedures as the prior study on new data, the reproducibility of the study is demonstrated.^{1,2} Does the current problem with reproducibility risk the good standing of the scientific enterprise?³ In a survey of 1576 researchers conducted by *Nature* in 2016, 90% agreed there was a crisis of reproducibility and just over half agreed that the crisis was significant.⁴ Reproducibility in science has been described as a “lynchpin of credibility,” and when credibility is lacking, both trust in science and the value of science declines.⁵ Reproducibility concerns have negatively affected confidence in some scientific disciplines⁶ and in certain political circles it has been used as fodder to further partisan policy aims and justify ideologically driven regulatory reforms.⁷ The stakes are high, and remedying the ongoing degradation in trust should be a top priority for scientists.⁷

A primary cause of the reproducibility crisis has been described as a failure to adhere to sound scientific practices coupled with pressures to publish or perish.⁸ Poor training in rigorous experimental design, research standards, and objective evaluation of data are among the key factors influencing the crisis.^{9,10} Although the National Institutes of Health (NIH) requires training in research ethics, institutions may not provide training in rigor and reproducibility of science. The NIH has called for a change in culture that includes both improved transparency as well as rigor and reproducibility.¹¹ This requires shifting values toward process as opposed to outcomes; focusing on research protocols, ethics, and quality of study design³ and reporting. Focusing on the process would encourage taking the time to assure high quality study designs. The NIH has reached out to the scientific community to take action.⁸ In response to this call, we obtained NIH funding via an administrative supplement to our Medical Scientist in Training Program for a curriculum project aimed at first-year medical students enrolled in research programs (MD/PhD and Clinical and Translational Research Pathway [CTRP]).

This paper describes the curriculum developed and our experiences with our first class of trainees.

METHODS

The study was approved by the University of Massachusetts Medical School Institutional Review Board. We conducted a

mixed-method evaluation. The protocols for evaluation were prespecified. For full transparency, we include all information related to the class development in the Tables and the Supplementary Materials.

Pedagogical approach

We designed the rigor and reproducibility curriculum to be flexible and to be responsive to the trainee’s intellectual interests, level of experience, and time available for learning. We define competencies as knowledge, skills, and abilities, and identified core competencies in rigor and reproducibility to guide our curriculum development. We believe that competency-based curricula provide clarity of the learning direction, stimulate accountability in the process of learning, and provide a framework for evaluation of learning.^{12,13} The University of Massachusetts Medical School switched to emergency remote educational activities in March 2020 using Zoom.

Trainees

All trainees were in their first year of medical school (1) MD/PhD, or (2) CTRP students. MD/PhD students complete a laboratory rotation before medical school. The MD/PhD program provides an integrated curriculum with emphasis on problem solving and small group learning and additional research coursework. The CTRP, a highly competitive program for medical students, offers research coursework and experiential learning in parallel with the medical school curriculum (Table S1). This was a required one credit class.

Topics and developing content for topics

Two course instructors defined the class parameters (eight, 2-h sessions, required preparation ~ 2–3 h per class). Class times were set by the medical school curriculum and were held once or twice a month between January and May 2020. We cast a wide net for existing resources, conducting multiple internet searches, reviewing existing publications, and teaching resources (NIH Rigor and Reproducibility Training Resources¹⁴; NIH National Institute of General Medical Sciences Clearinghouse for Training Modules to Enhance Data Reproducibility¹⁵; Life Science Teaching Resource Community¹⁶; and the Society for Neuroscience, Neuronline - Training

Modules to Enhance Data Reproducibility¹⁷). We assessed identified resources designed for graduate level learners (or higher). Relevant materials were sorted into preliminary topic areas. We combined or split these materials based on the volume of resources available and projected time for each topic.

We then focused on each class individually, performing a deeper dive into relevant publications and materials for each topic area using both Google and PubMed searches. Additional materials were identified and evaluated for appropriateness for inclusion as assigned or recommended preparation for class. Specific goals and objectives for each class were defined, and PowerPoint overviews with note pages were developed for instructor use. Trainees were assigned preparatory work, including required readings (1 or 2 articles from the literature) and/or podcasts, web-based learning modules, or video content. Optional resources were provided for those interested in deeper exploration. We then considered how to structure interactive classroom exercises to support classroom learning. Table 1 depicts the topics covered, the goals and learning objectives for each class, and the estimated time students will devote to preparation. From this extensive process, we finalized a list of competencies for the pre- and post-self-assessments. No curriculum approval process was required.

Interactive exercises

Table 2 summarizes the interactive exercises. Most require breaking out into small groups (4–5 students) for discussion (e.g., 15 min), with time for the groups to report back to the class with a summary of discussions. These activities were developed to build on the knowledge acquired from the required preparation. For example, in the first session, students were assigned one article.¹⁰ In their small groups, they were challenged to discover which of the NIH proposed ideas to improve rigor and reproducibility of research were implemented, identify new strategies developed since the article, and discuss why some ideas failed to be implemented. Discussion questions were provided. Next, groups selected a stakeholder perspective and discussed the pros and cons of each recommendation from their perspective. Through this discussion, trainees were challenged to think about the nuanced perspectives related to the topic. Examples provided showed the connection among didactics, theory, and practice. Before the class on quality control, trainees completed an assignment to (1) obtain a standard operating procedure, manual, or protocol from their research laboratory (or a classmate if they have not been assigned to a laboratory), (2) reviewed it, and (3) observed the practices in the laboratory. During in-class small group discussions, trainees shared what they learned from this exercise and discussed deviations from

protocols (if any) and how the laboratory could improve processes. If no deviations were observed, trainees were challenged to reflect on the reasons why.

Last, an entire class was devoted to two debates on open science frameworks—with half the class assigned to each debate topic. A debate is a formal method of presenting arguments in support of or against a given topic. Debates followed traditional formats, with time in breakout rooms to prepare rebuttal arguments. Students voted at the end via the polling option in Zoom.

Rigor and reproducibility project

The goals of this project were not to generate novel, innovative findings. Rather, the goals were twofold: (1) to provide additional insight into the importance of learning techniques and processes to improve the rigor and reproducibility of scientific research, and (2) to learn more about the importance of topics covered in the class to their own field (e.g., reporting requirements). Trainees learned with the backdrop of their own research questions in small groups (3–5 students). Each group selected one type of project:

1. Reproducibility project: Duplicate the results of a prior study using the same materials and procedures as the original investigator; trainees used the original data but applied their own analyses and interpretations.
2. Replication project: Duplicate the results of a prior study using the same procedures but with new data; determine generalizability to different subjects, age groups, racial/ethnic groups, locations, cultures, etc.
3. Blind data analysis: Apply techniques to obscure meaningful results, while showing enough of the data structure to deal with issues (e.g., outliers and confounders) and once all these issues are dealt with in the “altered data set,” rerun process on real data.

Trainees defined the project scope based on the time available. Depending on the difficulty of the project and challenges experienced, variation in the work completed by the end of the term was expected. Project groups selected the data available via one of the many public data resources. Trainees were cautioned that some data repositories have approval processes. We encouraged trainees to use open science tools or approaches to improve the transparency of their work. Trainees could elect to use one of the tools discussed (e.g., Experimental Design Assistant [EDA], open science framework, electronic laboratory notebook, or other methods). At four times during the course, trainees presented their progress to the class (see Table 2). The last class was devoted to final project presentations. Table 3 shows the topics, challenges, and insights gleaned from this assignment.

TABLE 1 Goals, learning objectives, estimated time commitment, and actual median and range of time on Blackboard Learning Management System (hours) stratified by module

Module title	Goals and learning objectives	Cumulative time Commitment ^a
Reproducibility crisis	<p>Goal: To introduce the origins and history of the Reproducibility Crisis.</p> <ul style="list-style-type: none"> Describe the origins of the reproducibility crisis. Know what the NIH response to the reproducibility crisis has been. List key stakeholders and describe strategies for addressing the reproducibility crisis. Define reproducibility, replication, and generalizability. 	<p>Expected: 4 h</p> <p>Median time on Blackboard: 1.2</p> <p>Range: 0–8.2</p> <p><5 min: 6%</p>
Evaluating rigor of prior research	<p>Goal: To define the requirements for an NIH scientific premise and provide basic skills to evaluate the rigor of existing research studies and proposals.</p> <ul style="list-style-type: none"> Describe the role and importance of rigor and reproducibility in NIH proposal writing and NIH scientific review. Describe the importance of scientific premise in NIH proposal preparation. Critique scientific premise statements. 	<p>Expected: 4 h</p> <p>Median: 0.8</p> <p>Range: 0–8.9</p> <p><5 min: 24%</p>
Rigorous experimental design and bias	<p>Goal: To review the elements of experimental design, tools and standards – including sex as a biological variable (NIH priority); to highlight areas of potential bias.</p> <ul style="list-style-type: none"> Discuss the importance of rigorous experimental design and documentation for transparency and replication. Describe when to include sex as a biological variable in research. Define bias and the sources of bias in the conduct of science. Assess bias using the Cochrane Collaboration’s tool for assessing risk of bias in randomized trials. Develop a prospective experimental design that comports with appropriate guidelines. 	<p>Expected: 4 h Median: 0.8</p> <p>Range: 0–6.1</p> <p><5 min: 47%</p>
Biological variables, authentication and QC	<p>Goal: To provide an overview of quality procedures for biomedical research, including authentication procedures. To provide an opportunity to discuss implementation challenges in laboratory settings.</p> <ul style="list-style-type: none"> Describe the key elements to include in an authentication plan for an NIH grant application. Describe quality practices important to basic biomedical research. Discuss the implementation of quality practices. 	<p>Expected: 5 h</p> <p>Median: 0.02</p> <p>Range: 0–2.6</p> <p><5 min: 65%</p>
Reporting expectations	<p>Goal: To review reporting guidelines used for manuscript preparation and to provide an overview of image processing and manipulation as it applies to clear and accurate reporting.</p> <ul style="list-style-type: none"> Describe how image data may be evaluated to determine whether manipulation has occurred. Describe software tools used to inspect images for manipulation. Using an article of your choosing, evaluate how well authors adhere to transparent reporting publication guidelines. 	<p>Expected: 4.25 h</p> <p>Median: 1.2</p> <p>Range: 0–5.0</p> <p><5 min: 18%</p>
Implementing transparency	<p>Goal: To present a workflow that promotes transparency including detailed record keeping and data management.</p> <ul style="list-style-type: none"> Describe the role of lab notebooks in promoting rigor and reproducibility. Describe the roles of the data management plan, metadata, and data dictionary. Describe the challenges and benefits of increased scientific transparency. Critically reflect on practices in your laboratory and consider possible steps toward increased transparency. 	<p>Expected: 3.5 h</p> <p>Median: 0.01</p> <p>Range: 0–1.8</p> <p><5 min: 88%</p>
Open science	<p>Goal: To provide an overview of the principles of open science and practical steps that can be undertaken to promote its implementation.</p> <ul style="list-style-type: none"> Define “open science.” Describe the overall goals of open science. Describe the challenges to the implementation of open science. Describe institutional changes that promote rigor and reproducibility. Select an open science objective and identify changes to current practices that promote its achievement. 	<p>Expected: 3.5 h</p> <p>Median: 0.4</p> <p>Range: 0–5.0</p> <p><5 min: 29%</p>
Total time on Blackboard Learning Management System across all elements of the class (hours): mean: 12.9, SD: 5.9		

Abbreviations: NIH, National Institutes of Health; QC, quality control.

^aIncluding assigned readings, preparatory work, and ongoing work on the project.

TABLE 2 Interactive in-class exercises

Topic	In-class activities
Reproducibility Crisis 2 small group (3–5 students) discussions 15 min each 5-min summary of each group's discussion	<p>Discussion 1: In the 5 y since the Collins and Tabak article, which of the proposed ideas for NIH to address the reproducibility crisis have been implemented? Conduct some internet “sleuthing.” Your team may provide a general scan or a “deep dive” of one aspect. Why do you think some ideas succeeded and others failed? What are your thoughts about the potential impact for the implemented changes to address the reproducibility issues? During the course of your internet searching, did you come across any new ideas that have been implemented by NIH (or others) to address the reproducibility crisis?</p> <p>Discussion 2: Each team select one of the stakeholder roles (i.e., student, journal editor, academic institution (e.g., promotion committees), funder, researcher). From your stakeholder perspective, discuss strategies to address the reproducibility crisis.</p> <p>Consider what is already being done (or a new idea!) and how “success” of the strategies might be measured. Discuss the implications for implementing (pros/cons) from your stakeholder perspective. List pros/cons from other stakeholder perspectives.</p>
Evaluating the rigor or previous research, scientific premise 1 small group (3–5 students) discussion 30 min for discussion with 5-min summary from each group	<p>Each group discusses the high-level overview of an F30 proposal assigned to the group. Based on readings regarding the importance of scientific premise in NIH review of proposals, what specific prior research studies would your group like to see referenced in support of the scientific premise of this NIH proposal? Has the research your group believes is necessary been done? What is the quality or the previous research which forms as the foundation for the current proposal? Discuss how to determine the rigor of the studies you would like to see before you would highly score the application.</p> <p>Project presentation 1: Students present reproducibility or replication project, why selected, team members, outline of what the team believes will be reasonable to accomplish (e.g., download data, recreate the sample, recode variables, run preliminary analyses)</p>
Rigorous experimental design and bias	<p>Class watches NIH Video together (Module 2: Blinding and randomization³⁰) Followed by small group discussion with questions provided by NIH (e.g., can you think of a particular instance in which blinding and randomization could have a dramatic impact on the results?) Cochrane assessment bias tool exercise. Hands on exercise with Experimental Design Assistant Tool³¹</p>
Biological variables, authentication, and quality control	<p>Discussion 1: Class watches NIH video together (Module 4: Sample size, outliers, sex as a biological variable³⁰) Followed by small group discussion with questions provided by the NIH (e.g., Have you or someone you know only used male mice in an experiment as a way of avoiding the “sex issue?” Do you think this is appropriate? Does it depend on the type of experiment being done?).</p> <p>Discussion 2: Before the class on quality control, trainees complete an assignment to (1) obtain a standard operating procedure, manual, or protocol from their research laboratory (or a classmate if they have not been assigned to a laboratory), (2) review it, and (3) observe the practices in the laboratory. During in-class small group discussions, trainees shared what they learned from this exercise and discuss deviations from protocols (if any) and how the laboratory could improve processes. If no deviations were observed, trainees were challenged to reflect on reasons why.</p> <p>Project presentation 2: Recap of topic, overview of methods</p>
Reporting expectations 1 small group (3–5 students) discussion 20 min for discussion with 5-min summary from each group	<p>Each small group assigned an article. <i>Who is to blame?</i> Summarize evidence in support of (researchers, sponsors, editors) based on the article assigned to your group. <i>What can be done about it?</i> Brainstorm ideas to address publication bias given your thoughts and the evidence regarding who is “to blame.”</p> <p>Project presentation 3: Recap of topic, tasks accomplished, challenges experienced, and preliminary results</p>
Implementing transparency 2 small group (3–5 students) discussions 15 min each 5-min summary of each group's discussion	<p>Class watches NIH video #1 (Module 1: Lack of transparency³⁰) Followed by small group discussion with questions provided by NIH (e.g., Do you think the corresponding author should have handled the situation differently?).</p> <p>Moving Forward: Individually, critically reflect on practices in your laboratory and consider possible steps toward increased transparency. What are the most pressing needs for improving practices in your laboratory? How would you address them moving forward?</p>
Open science	<p>Standard debate format (see text). Debate 1: Should scientists at our institution be required to use an open science framework for their research? Debate 2: Should federal funders of research in the United States (e.g., NIH, NSF, etc.) participate in Plan S?</p>
Reproducibility/ replication projects	<p>Project presentation 4: Team, topic, methods, open science / transparency methods used, challenges, preliminary results, unexpected aspects of the project, findings, transparency, thoughts on open science, transparency, rigor, etc.</p>

Abbreviations: NIH, National Institutes of Health; NSF, National Science Foundation.

TABLE 3 Challenges and insights from final team projects (3–5 students per team)

Title	Challenges	Insights
TCGABiolinks: An R-based, Open Source Tool for Genomic Analysis of Published TCGA data ²⁶	Updates to TCGABiolinks were not backward compatible. Initial release of software was in 2015. Modifications were required that prevented exact replication.	Exciting to see how much data set has grown since 2019 publication. Exciting to be able to replicate findings. Relatively stress-free experience because of excellent documentation.
Meta-analysis of antidepressant efficacy ²⁷	Study transparency was overall quite good. Data set was available on-line and well-documented.	Challenged by calculation of metrics (e.g., credible interval)
Association of electronic cigarette use with subsequent initiation of tobacco cigarettes in US youths ²⁸	Figuring out what data were used. Inability to replicate the sample because variables to define inclusion/exclusion criteria were not available in the public data set. Figuring out what weights were used. Lack of detail prevented ability to replicate the recoding.	Publicly available data sets may lack PHI needed to replicate samples. Independent studies using data from national studies may not publish their own data extract. Replication was impossible.
Re-examination of data: EGFR as receptor of interest on monocytes, causal determination of HCMV on EGFR ²⁹	No raw images were included in the omics di repository. Authors made data available, but files were too large to process in R Studio; work arounds identified, but package no longer available with latest version of R. Details provided about wet laboratory procedures certain biological descriptions were ambiguous, but nothing about the data cleaning, missing data, statistical techniques used, and testing of assumptions. No response to emails sent to the authors for more information.	Data access issues and technical challenges were surprising (backward compatibility). Evidence of image compression artifacts, value inversions, narrow cropping; such issues may be a pervasive issue in biological sciences. Need to include data for all components of a study with user-friendly documentation. The importance of sharing scripts for data cleaning and statistical practices.

Abbreviations: HCMV, human cytomegalovirus; PHI, protected health information; TCGA, The Cancer Genome Atlas.

Evaluation

The class was graded as pass/fail. No examinations were given, but students received feedback from professors regarding their work on the group projects and debates. For the evaluation of the curriculum, we modified our institutional review board (IRB) protocol to conduct focus groups over Zoom because of coronavirus disease 2019 (COVID-19). One author (C.D.), an experienced qualitative researcher and one of the designers of the curriculum, conducted two interviews with the course directors (2 women; 30 and 40 min each) and four student focus groups using a convenience sample (volunteers: 7 men and 6 women) recruited via email. Semistructured interview guides were used (Tables S2,S3). Focus groups were recorded via Zoom, video-teleconferencing with a digital recorder as back-up (average length: 66 min; range: 45–85) and a research assistant taking notes in two focus groups. Audio was professionally transcribed. Participants did not review transcripts or findings. We used simple thematic analysis (themes derived from the data) and coded by C.D. using NVivo software.¹⁸ We achieved saturation, as many comments were repeated in later focus groups. Participants received a \$10 gift card. The authors also observed the debates and project presentations. The interviewer used techniques designed to elicit both positive and corrective feedback. She was unfamiliar with the students and one instructor. Participants knew that C.D. is faculty and had a role

in designing the course. Students conducted pre- and post-self-assessments of competencies (Table S4). Quantitative analyses included descriptive statistics (means and SDs) and paired *t*-tests with *p* values less than 0.05 considered statistically significant.

RESULTS

Table S1 shows that 44.4% of class participants were women and 16.7% were from racial/ethnic groups under-represented in the biomedical sciences. Grade point averages (GPAs), Science GPA, and Medical College Admission Test (MCAT) scores were similar between the student groups. Table 1 shows that the average time on Blackboard was less than the expected preparation time. Because materials were meant to be downloaded rather than read online, we could not estimate the actual student preparation time.

Table 4 shows that overall, the course and its components were very favorably reviewed by both students and instructors via data collected in focus groups and in-depth interviews. Among students, course content was seen as novel, important, relevant, and practical—and teaching strategies were generally seen as effective. Students particularly appreciated short lectures and interactive small group exercises. A review to reduce redundancy was requested. Preparatory assignments in a variety of formats

TABLE 4 Student focus group and instructor interview findings

Category	Strengths	Suggestions
<i>Student comments</i>		
Overall	<p>Dedicated time to think about and discuss rigor and reproducibility, learning from others' experiences, opportunity to meet new people; cohort effect a "huge bonus."</p> <p>Course organization was effective: from overview to different components "and each time coming around with how can we do this better?"</p> <p>"For people who do not know much about reproducibility, in 3 months, I thought it was incredible."</p>	<p>Review course to reduce repetition: "Sometimes it got a little repetitive... we started doing similar things as we got closer to the end of the class."</p> <p>Timing: (1) Offer in the Fall semester as an introduction to research training, (2) stretch the class out over the course of an academic year to integrate with other teaching, (3) run concurrently with a laboratory rotation.</p> <p>Have a teaching assistant to develop summaries of preparation work for each class, update the website, and assist with final projects.</p>
Content	<p>Reviewed general concepts and provided a method of thinking; "a different way... [to] look at things like open science and transparency"; Important principles were covered relevant to future careers; "keep them all."</p> <p>Addressed NIH expectations; Focused on practical tools/available resources; best practices for maintaining laboratory notebooks was helpful/useful; Cochrane Guidelines session was valuable; "I enjoyed the topics that we were taught," Provided concrete examples of good and bad science.</p>	<p>Delve deeper into what makes good research—like how to set up an RCT, how to make figures attractive in a paper or abstract, or graphical abstracts.</p> <p>Focus more on best practices.</p> <p>Include more good and bad examples.</p> <p>Include more on bioinformatics and database research.</p> <p>More analysis of mistakes/misconduct of others.</p>
Lectures	<p>Lectures were short and to the point (first lecture was most helpful/effective); defined and clarified terms, explained concepts; "Didn't really dive too deeply into the weeds"; Image falsification/analysis lecture was particularly interesting; Chat box for questions worked well.</p>	<p>Reduce time in lectures to the bare minimum; make them more interactive (e.g., quiz format); lectures sometimes "blended together"; provide "coming attractions" for next class—stress essential preparation (for in-class exercises); consistently explain concepts and then show an example; add guest speakers with expertise in the area.</p>
Small group exercises	<p>Exercises effectively applied concepts from the lecture; interactive in nature; evaluating and critiquing specific papers was valuable; class presentations allowed for peer teaching; appreciated the opportunity to learn from peers through presentations and discussions; effectively promoted engagement (everyone had a say on a topic).</p>	<p>Discussion time was sometimes too short; devote more time to small group work; first exercise was on an unfamiliar topic (acupuncture); replace with more familiar content; taking a study and formulating a replication plan was too much for a short in-class exercise; redesign to make it more feasible; group reports could be repetitive when each group was tasked with the same thing.</p>
Preparatory assignments	<p>Good to have a mix of assignment types—engaging.</p> <p>Video assignments were valuable "it's a nice break and really good to just let it soak in."</p> <p>Podcasts were a welcomed alternative to articles; "appreciate a more entertainment accessible source material."</p>	<p>Reduce/consolidate the number of readings "We just don't have the time to do it."; for webinars, to be able to speed up the playback or have transcript; clarify purpose of each reading; provide a distilled summary for prep work; add readings that reflect clinical relevance.</p> <p>Provide in this format: (1) summary document of all key points; (2) essential preparation (discussed in class); (3) required preparation; (4) suggested/ recommended preparation; (5) additional resources.</p>
Quizzes	<p>Quick and not a burden; "pretty straight-forward"; helpful to get a "gist" of the most important take-aways; seemed helpful to instructors to know what students absorbed.</p>	<p>"Sometimes the quizzes were a bit of a head scratcher"—make them more relevant to key objectives and reinforce main points; build in reminders to ensure completion of quizzes.</p>
Final project	<p>A valuable exercise and longitudinal experience; different options were available for type of project meeting different student needs; effectively synthesized learning; led to impressive efforts and presentations by peers—"it seemed like we were experts in what we were doing."</p>	<p>Clarify the goals of the final project; show an example; provide a more definitive guide on how to do the project; provide a "how to" manual for finding articles with an available dataset; have more frequent meetings with instructors for advice and guidance; add option for a proposal for a replication study instead.</p>

(Continues)

TABLE 4 (Continued)

Category	Strengths	Suggestions
<i>Instructor comments</i>		
Overall	<p>“I thought it was a phenomenal course.”</p> <p>“This is going to go better than you think, and so prepare for a really good product at the end and do something with it. Like, leverage what the students are producing—capture it in some way and... you can use some of these things on websites. You can use it for recruiting. There’s a lot of product that’s going to come out of this that you want to leverage and make some time to evaluate it and use it.”</p>	<p>“I think all the [medical] students should get this [course]... you could debate whether the first year is the place to do it or later... every student should be exposed to it.”</p> <p>[Integrate breaking news] “Fortuitously, on the first day of class was the same day that a Nobel Laureate in chemistry withdrew an article from Science because it wasn’t replicable...”</p> <p>“Increase a little bit the amount of primary research material... for example, how journal policies have evolved... a deeper dive into the problem.”</p>
Content	<p>“The content was excellent... I wouldn’t get rid of any of the content... it touched on all the points that are relevant... from the ethical to the very technical.”</p>	<p>“One way to grab a medical student or a nursing student’s attention is to give examples that are clinically relevant... have some readings or some examples where the irreproducibility of the research resulted in an adverse clinical outcome.”</p>
Lectures	<p>Content was good and covered necessary topics.</p>	<p>“Make the lectures a bit shorter”... add content and remove slides like Goals & Objectives.”</p> <p>“More information about quantitative aspects... particularly epidemiology.”</p> <p>“Bring in an outside speaker... who’s got more expertise... who’s a real expert in, for example, image manipulation.”</p> <p>“Invited speakers... people that wrote the stuff.”</p>
Small group exercises	<p>“Student engagement was very high... the ability to engage all of the students all of the time was a particular strength.”</p>	<p>After small-group exercises and reporting by small groups: “Maybe something to tie it all together... has your mindset changed?... maybe just [add] like a bit of a summary or final thoughts... go back to the exact same [original] question... was your initial knee-jerk reaction correct or not? What did you learn that now would make you think something was different?”</p>
Preparatory	<p>“One of the great successes is that they [first-year medical students] didn’t really have to prepare—and there was no penalty. And they were 100% during that time... they get a lot out of it.”</p>	<p>Move some of the content of the lectures to preparatory work.</p>
Final project	<p>“The [final] project...that was outstanding...” “I thought it was the best part.”</p>	<p>Use bioRxiv https://www.biorxiv.org/: “Maybe that would be the idea, to pick papers that they would reanalyze on bioRxiv... all the data is there.”</p> <p>Final Project: “Give them the project with an eye toward, you’re going to publish this, or at least you are going to blog it... almost all they need to do is narrate their presentation... it would not be too much work if the goal was, okay, now you are going to post it.”</p>

Abbreviations: NIH, National Institute of Health; RCT, randomized controlled trial.

were appreciated, however, most students expressed concern about lack of time for course preparation. A reduction of reading assignments, sharper focus, and streamlining of preparation work may be required. The majority did not take the 5-min quizzes, despite encouragement from course leaders and extended deadlines. Quizzes had no bearing on their grade. For those who did complete the weekly quizzes, the scores were poor despite a low level of difficulty (data not shown). In the focus groups, students noted that quizzes

were helpful, but in need of some revision. The final project was considered valuable and flexible resulting in impressive final presentations. Students offered a wide range of specific suggestions for course improvement. Instructors’ comments were enthusiastic in terms of content, student engagement, and overall impact. Instructors’ key suggestions for improvement focused on fine-tuning the course through shifting content, adding new content, refining some methods, and adding resources, such as guest speakers.

TABLE 5 Self-assessments^a of competencies, before and after rigor and reproducibility class

	Before (<i>n</i> = 18) ^b	Before (<i>n</i> = 11) ^b	After (<i>n</i> = 11) ^b
1. The origins of the reproducibility crisis.	1.72 (1.32)	2.36 (1.29)	5.45 (1.29)
2. Strategies for addressing the reproducibility crisis.	2.67 (0.97)	2.73 (0.90)	5.82 (1.08)
3. The NIH response to the reproducibility crisis.	2.44 (1.25)	2.45 (1.13)	5.45 (1.51)
4. The role and importance of rigor and reproducibility in NIH proposal writing and scientific review.	3.78 (1.26)	3.45 (1.04)	6.09 (0.94)
5. The importance of scientific premise in NIH proposal preparation.	3.50 (1.15)	3.54 (1.04)	6.00 (0.94)
6. Critically assess sample scientific premise statements.	3.39 (1.50)	3.45 (1.75)	5.36 (1.63)
7. The importance of rigorous experimental design and documentation for transparency.	4.78 (1.17)	4.73 (0.65)	6.45 (5.20)
8. The importance of including sex as a biological variable in research.	4.50 (1.29)	4.27 (1.10)	6.50 (0.71)
9. Bias and the sources of bias in the conduct of science.	4.56 (1.04)	4.72 (1.19)	6.00 (1.00)
10. Assessing bias using the Cochrane Collaboration's tool for assessing risk of bias in randomized trials.	1.28 (0.67)	1.27 (0.65)	4.64 (1.63)
11. Developing a prospective experimental design that comports with appropriate guidelines.	3.39 (1.09)	3.55 (1.04)	5.45 (1.29)
12. Key elements to include in an authentication plan for an NIH grant application.	1.72 (0.96)	1.91 (1.04)	4.73 (1.68)
13. Quality practices important to basic biomedical research.	4.06 (1.21)	4.00 (1.00)	6.18 (0.98)
14. Implementation of quality practices for basic biological research.	3.83 (1.42)	3.73 (1.27)	5.82 (0.98)
15. Evaluation of image data to determine whether unacceptable manipulation has occurred.	2.44 (1.62)	2.82 (1.78)	5.64 (1.12)
16. Software tools used to inspect images for manipulation.	1.89 (1.13)	2.18 (1.25)	4.82 (1.25)
17. Evaluating adherence to transparent reporting publication guidelines.	2.39 (1.04)	2.45 (1.13)	5.40 (1.17)
18. The role of laboratory notebooks in promoting rigor and reproducibility and transparency.	4.56 (1.58)	4.82 (1.60)	6.45 (0.93)
19. The roles of the data management plan, metadata, and data dictionary in promoting reproducibility and transparency.	3.56 (1.95)	3.72 (2.10)	5.73 (1.42)
20. Challenges and benefits of increased scientific transparency.	3.94 (1.43)	4.09 (1.45)	6.18 (0.75)
21. Critically assessing practices in your laboratory and consider possible steps toward increased transparency.	3.56 (1.82)	3.45 (1.69)	5.82 (0.75)
22. "Open Science" and its overall goals.	2.94 (1.55)	2.91 (1.64)	6.18 (0.98)
23. The challenges to the implementation of Open Science.	2.39 (1.20)	2.45 (1.21)	6.09 (1.04)
24. Identifying changes to current practices that promote Open Science.	2.67 (1.41)	2.73 (1.49)	5.73 (0.79)
25. Institutional changes that promote rigor and reproducibility.	3.17 (1.20)	3.55 (0.93)	5.64 (0.81)

Abbreviation: NIH, National Institutes of Health.

^aStudents ranked each item on a scale where 1 = know nothing, 2 = very basic understanding, 3 = low/moderate understanding, 4 = moderate understanding, 5 = high/moderate understanding, 6 = strong understanding, and 7 = highly competent.

^bAll students completed the assessment before class, 11 students completed the post-assessment. All paired *t*-tests were <0.05 for the all students and MD/PhD students (*n* = 7), pre-post scores were not statistically different for Clinical and Translational Research Pathway students (*n* = 4) 5, 6, 9, 10, 11, 12, 16, 17, 19, 22, and 25.

All students completed the pre-course self-assessment of competencies (Table 5) The average score for 11 of the 25 competencies was between 1.0 (know nothing) and 2.9 (very basic understanding). The average score for 9 competencies was 3.0 to 3.9 (low/moderate understanding), and the average for the remaining 5 competencies was 4.0 to 4.9 (moderate understanding). For all students who completed the pre- and post-assessments (*n* = 11), self-reported competency increased (*p* values <0.05). All were improved

among the MD/PhD students, but pre- and post-scores were similar on 11 items for the CTRP students (with 4 completing).

DISCUSSION

Using a highly interactive, "flipped" classroom pedagogy, we have demonstrated that first-year medical and MD/PhD

students can improve their competency in scientific rigor and reproducibility. The students in the class conducted research across the clinical and translational research spectrum, although there were more basic scientists than clinical researchers. Our mixed method evaluation provided evidence of enthusiasm for the course materials from instructors and students, effectiveness of the curriculum, and areas for improvement.

We designed the curriculum with the goal of training reflective practitioners skilled in both knowledge and ways of thinking about rigor and reproducibility across the translational research spectrum. Although two research ethics courses are currently offered at the University of Massachusetts Medical School, neither addresses the core competencies covered in the rigor and reproducibility class, neither uses debate as a learning experience, and neither had students work in teams on projects. We believe the curriculum is well-suited for all graduate students. We developed educational modules using a reflective learning framework.^{19,20} Trainees were provided opportunities to talk, listen, read, write, and reflect as they approached content through problem-solving exercises in small groups, simulations, and case studies, all of which require trainees to apply what they are learning.²¹ We used the reflective practice approach because we thought it would help medical students hone transformative (double loop) learning skills.^{12,22} Although the reliance on a static frame of reference (“single loop learning”²³) meets the professional needs when theory and knowledge are constant and challenges/dilemmas are predictable, it falls short in clinical and translational research because theory and knowledge are dynamic and challenges facing the field are unpredictable.²³ In this class, students did not see the same problems twice. Exercises allowed them to apply knowledge in one field to the current problem at hand. The exercises also required reflection on each topic. This learning paradigm allowed the trainees to improve their ability to critically analyze a problem based on experience, knowledge, critical thinking, and intuitive knowledge developed through previous reflections.²⁴

Each student participated in a scientific debate, a collaborative learning exercise, which provided the students with an opportunity to practice scientific argumentation. A social process in which trainees build, question, and critique claims using evidence,²⁵ debates provide opportunities for students to hone the 4 elements of scientific argumentation: (1) evidence (use of high-quality evidence), (2) reasoning (articulate how evidence supports claims), (3) social interaction (build off other’s ideas), and (4) competing claims (critique and alternative explanations). The debaters presented their reasons and evidence to persuade the rest of the class. Participants sharpened their thinking and speaking skills through preparation and participation in the formal debate. The trainees not only learned more about the topic, but also had opportunities

to further develop persuasive speech skills, increase collaboration skills, and apply conflict resolution abilities.

Last, we wanted students to have hands-on experience with replication, reproducibility, or blind data analytic techniques, as we believed that some lessons can only be learned by doing. With the required rigor and reproducibility project, trainees gained skills in “problem setting”—naming the things to learn and framing the context in which they learn.¹³ The project was viewed as an outstanding experience for instructors and students. Students reflected that having more time to meet individually with their instructors to discuss the projects or having a teaching assistant to assist them with their projects would have been helpful. Despite these challenges, the instructors, peers, and faculty present (K.L. and C.D.) during the final presentations of the projects were impressed with the clear level of competence achieved via the project.

Our findings represent the experiences of one class. We did not know whether students had rigor and reproducibility training before entering this class, nor did we evaluate the extent to which the class worked equally well for basic scientists and clinical researchers. We unexpectedly learned that the highly interactive, small group classes worked well with breakout rooms in Zoom. We used the polling option to vote after the debates. Quizzes should be replaced with end of class polls to reinforce key messages from each session and provide opportunity for a brief discussion and clarification if students are unable to accurately answer the questions. We suggest instructors send out reminders before class so that students will prepare in advance. We encourage instructors to make time during the last class to complete self-assessments. Plans to modify the curriculum based on the evaluation results herein before offering it to other groups of students at our university are in process.

CONCLUSION

Formal training is needed to raise awareness of the reproducibility crisis and improve the rigor of research conducted. Highly interactive exercises coupled with a hands-on replication group project provided a pathway for students to gain competencies in improving the rigor and reproducibility of scientific research. Despite the limited time available to complete assigned preparatory work given the other demands on first-year medical students’ time, the flipped classroom pedagogy appeared to be successful. Rigor and reproducibility can be taught in a highly interactive format and in a remote format and doing so results in improved rigor and reproducibility competence.

ACKNOWLEDGMENTS

The authors wish to thank the instructors and students in the University of Massachusetts Medical School, MDP740 Spring 2020.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

K.L.L. and C.E.D. wrote the manuscript. K.L.L. and C.E.D. designed the research. K.L.L. and C.E.D. performed the research. K.L.L. and C.E.D. analyzed the data.

REFERENCES

- Bollen K, Cacioppo J, Kaplan R, Krosnick J, Olds J. Social, behavioral, and economic sciences perspectives on robust and reliable science. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. 2015 <https://www.nsf.gov/sce/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf>. Accessed December 4, 2020.
- Pellizzari E, Lohr K, Blatecky A, Creel D. *A Primer on Semantics and Implications for Research*. Research Triangle Park, NC: RTI Press. 2017 <<https://www.rti.org/rti-press-publication/reproducibility-primer/fulltext.pdf>>. Accessed December 4, 2020.
- Does the Public Trust Science? Trust and Confidence at the Intersections of the Life Sciences and Society. A Workshop Summary. National Academies of Sciences, Engineering, and Medicine. 2015 <<https://www.nap.edu/download21798>>. Accessed December 4, 2020.
- Baker M. Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help. *Nature*. 2016;533:452-454.
- Marcus E. Credibility and reproducibility. *Cancer Cell*. 2014;26(6):771-772.
- Anvari F, Lakens D. The replicability crisis and public trust in psychological science. *Compr Results Soc Psychol*. 2018;3:266-286.
- Salteli A. Why science's crisis should not become a political battling ground. *Futures*. 2018;104:85-90.
- Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res*. 2015;116:116-126.
- Carp J. The secret lives of experiments: methods reporting in the fMRI literature. *NeuroImage*. 2012;63(1):289-300.
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014;505:612-613.
- Samsa G, Samsa L. A guide to reproducibility in preclinical research. *Acad Med*. 2019;94:47-52.
- Smith S, Dollase R. AMEE Guide No. 14. Outcome-Based Education: Part 2--planning, implementing and evaluating a competency-based curriculum. *Med Teach*. 1999;21:15-22.
- McGaghie W, Miller GE, Sajid AW, Telder TV. Competency-based curriculum development on medical education: an introduction. *Public Health Papers*. 1978;68:11-91.
- NIH. Rigor and Reproducibility <<https://www.nih.gov/research-training/rigor-reproducibility/training>>. Accessed February 1, 2020.
- NIH. Clearinghouse for Training Modules to Enhance Data Reproducibility <<https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx>>. Updated May 20, 2020. Accessed February 1, 2020.
- Bennett C, Matyas ML III. Reporting Results: Controls in Animal Studies for Rigor and Reproducibility (Digital Presentation)

<<https://www.lifescitrc.org/resource.cfm?submissionID=11968>>. Published 2019. Accessed February 1, 2020.

- Neuroline. Promoting Awareness and Knowledge to Enhance Scientific Rigor in Neuroscience <<https://neuroline.sfn.org/collecion/promoting-awareness-and-knowledge-to-enhance-scientific-rigor-in-neuroscience>>. Accessed February 1, 2020.
- QSR NVivo. Melbourne, Australia 2000.
- Mezirow J. *Transformative Dimensions of Adult Learning*. USA: Jossey-Bass; 1991.
- Schon D. *The Reflective Practitioner: How Professionals Think in Action*. Avebury: Ashgate Publishing Ltd.; 1991.
- Meyers C, Jones T. *Promoting Active Learning: Strategies for the College Classroom*. San Francisco, CA: Jossey-Bass; 1993.
- Argyris C, Schon D. *Theory in Practice: Increasing Professional Effectiveness*. San Francisco, CA: Jossey-Bass Inc., Published 1974.
- Kolb D. *Experiential Learning: Experience As The Source Of Learning And Development*. Upper Saddle River, NJ: Prentice-Hall; 1983.
- Boud D, Walker D. *Experience and Learning Reflection at Work*. Melbourne, Australia: Deakin University; 1991.
- Osborne J, MacPherson A, Patterson A, Szu E. *Perspectives on Scientific Argumentation: Theory, Practice, and Research*, M.S. Khine (Ed.). New York, NY: Springer; 2012.
- Mounir M, Lucchetta M, Silva T, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLoS Comput Biol*. 2019;15(3):e1006701.
- Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357-1366.
- Berry KM, Fetterman JL, Benjamin EJ, et al. Association of electronic cigarette use with subsequent initiation of tobacco cigarettes in US youths. *JAMA Network Open*. 2019;2(2):e187794.
- Chan G, Nogalski MT, Yurochko AD. Activation of EGFR on monocytes is required for human cytomegalovirus entry and mediates cellular motility. *Proc Natl Acad Sci U S A*. 2009;106(52):22369-22374
- NIH. Rigor and Reproducibility <<https://www.nih.gov/research-training/rigor-reproducibility/training>>. Accessed February 1, 2020.
- The Experimental Design Assistant - EDA <<https://www.nc3rs.org.uk/experimental-design-assistant-eda>>. Accessed December 4, 2020.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Lapane KL, Dube CE. Rigor and reproducibility training for first year medical students in research pathways. *Clin Transl Sci*. 2021;14:1026-1036. <https://doi.org/10.1111/cts.12966>