

## Measuring the impact of conceptual inquiry-based labs

Danny Doucette

*Department of Physics, North Carolina State University, Raleigh, North Carolina, 27607, USA*

Conceptual inquiry-based introductory physics labs deploy PER-informed pedagogical strategies in physics labs with the aim of helping to improve students' understanding of physics concepts. Unlike traditional labs (which tend to be highly-structured and focus on verification of scientific relationships via precision measurements) and skills-based labs (which tend to eschew the aim of helping students learn physics concepts in favor of teaching experimental skills), some studies have suggested that conceptual inquiry-based labs may have a positive impact on student conceptual understanding, as measured by concept inventories. This paper reports on a randomized controlled study that compares student conceptual gains in electricity and magnetism between a conceptual inquiry-based lab and a skills-based lab. Following a difference-in-differences analytical strategy and using hierarchical linear modeling, the result from this study is that the conceptual inquiry-based lab provides no additional benefit to students' conceptual learning gains compared with the skills-based lab. Studies such as these may help physics departments make decisions about the goals and scope of transformations to their introductory lab courses.

## I. INTRODUCTION

Over several generations, a variety of theoretical [1–4] and statistical [5–9] analyses have raised doubts about whether traditional physics labs can positively impact students’ understanding of physics concepts. By the middle of the 20th century, a variety of colleges were moving toward skills-based (as they are called today [10]) introductory physics labs [11]. Then, following work by physics education research (PER) scholars starting around the 1990s, a new variety of introductory college physics lab started to emerge: the conceptual inquiry-based lab [12]. Today, as many physics departments re-evaluate their introductory lab courses, there is a need for research-based guidance: can PER-informed conceptual inquiry-based labs support students’ learning of physics concepts?

Traditional labs are typically described as highly-structured [13], with step-by-step instructions designed to help students efficiently make high-precision measurements using sometimes-unfamiliar apparatus, and with learning goals typically focused on “illustrat[ing] lecture courses” [14] and “verify[ing] or confirming laws, relations, [and] regularities asserted in text, class, or lecture” [12].

Like traditional labs, PER-informed conceptual inquiry-based labs aim to support conceptual learning for students. However, unlike traditional labs, conceptual inquiry-based labs tend to have less focus on step-by-step instructions for operating apparatus. Instead, they typically employ experiments that allow students to directly measure the relevant quantities, using familiar apparatus where possible, and focus on uncovering, rather than verifying, relations between variables. Instead of illustrating and verifying, conceptual inquiry-based labs aim to “help students acquire an understanding of a set of related physics concepts” [15].

Several investigations have attempted to measure the impact of conceptual inquiry-based labs on student understanding of physics concepts using concept inventories. However, such studies are difficult because students are often enrolled in a lecture course at the same time as the lab, and one may expect to see learning gains because of their enrollment in the lecture class. Therefore, these studies typically use a traditional lab as a control group. In one such study, students in a guided-inquiry lab achieved a larger pre-to-post gain in Force Concept Inventory (FCI) scores compared with peers in a traditional lab [16]. Another study found that students in guided inquiry labs outperformed peers in either traditional or open-ended labs [17]. A third study, comparing an inquiry-based lab with a traditional lab, found no difference on the FCI but a significant difference on the Mechanics Baseline Test (MBT) [18]. A fourth study, comparing inquiry-based and traditional labs, found no difference on either the FCI or MBT, but a significant difference on course grades [19]. Finally, a fifth investigation found a significant increase in FCI scores for students taking a conceptual inquiry lab course based on RealTime Physics [15] compared with a traditional curriculum [20].

These mixed results suggest that in some but not all cases conceptual inquiry-based labs can be effective at helping students learn physics concepts. Therefore, further work is needed to identify the features of conceptual inquiry-based labs that are effective in helping students to learn, how these mechanisms may vary between different types of institutions, and what range of outcomes are possible with different types of conceptual inquiry-based labs.

This paper contributes to addressing these concerns by reporting on a randomized controlled trial that compares a PER-informed conceptual inquiry-based lab with a skills-based lab using student scores on the Brief Electricity and Magnetism Assessment (BEMA) [21]. To my knowledge, this is the first controlled study that compares student outcomes on a concept inventory when comparing conceptual inquiry-based labs with skills-based labs.

## II. METHODOLOGY

This study considers matched pre/post BEMA scores from 263 students (159 in the conceptual inquiry-based lab, 104 in the skills-based lab; 77% participation rate) who were enrolled in a 1-credit, calculus-based lab course designed to accompany a 3-credit lecture course focused on electricity and magnetism. Students in this class typically pursue studies in engineering or the physical sciences at this large, land-grant university in the South of the USA. All of the students in the study were simultaneously enrolled in the lecture-based class that also addressed electricity and magnetism.

Students attended this lab class once per week for 2 hours. They worked in groups of 2 (or 3, if there was an odd number of students) to complete their in-person lab-work. Enrollment for sections of this lab was capped at 24 students, as there are 12 lab benches in the lab room. Supervision of the lab was provided by six graduate student teaching assistants (TAs), who received training, prepared equipment, and practiced doing the lab-work themselves at Friday lab meetings. Each TA taught 2 to 5 lab sections, depending on their teaching load and assignment for the semester. This semester, because of challenges associated with the return to in-person instruction, students completed 9 weeks of labs.

Randomization of the two lab curricula occurred via dice roll at the TA level during the first Friday meeting of the semester. Students were informed that there were two different lab curricula being run in the different sections, but were not able to switch lab sections (nor did any ask to).

Both lab curricula followed the same structure. Students were assigned to read the lab instructions before their lab. Individually, they completed a 5-item multiple-choice pre-lab quiz to assess their understanding of the instructions. During the 2-hour lab period, the students worked with their lab partner to complete the lab, following the lab instructions. They submitted a joint lab report by the end of the lab period. The lab report was either handwritten on paper and handed in, or written using a word processor and uploaded as a PDF doc-

ument to the learning management system, depending on the TA's preference. The lab report was graded using a generic rubric by the TA, and returned to the students. The rubric had two categories (completeness and correctness), each of which was evaluated with a score between 0 and 5. During the first few Friday meetings, the TAs graded some lab reports together in order to standardize their understanding of the rubric. Following the lab, but before the end of the day on Saturday, students completed a 5-item multiple-choice post-lab quiz to assess their understanding of the concepts and skills they ought to have learned during the lab.

The conceptual inquiry-based labs used lab investigations designed for the *Matter & Interactions*, 4th edition, textbook [22] (available at [instructormi4e.org](http://instructormi4e.org)). This is the textbook that was used in the students' lecture classes. Nine of the twelve lab experiments developed for Volume 2 (Electric and Magnetic Interactions) of the textbook were used in this course. Two small changes were made to the lab instructions: First, the recommended time for each lab was removed, in order to avoid giving the students the impression that they needed to rush while completing this lab-work. Second, numbers were added to each prompt in the lab assignment for which students were asked to calculate an answer, make a prediction, draw a diagram, etc., in order to provide some structure for the students' lab report.

The pre-lab and post-lab quizzes for the conceptual inquiry-based labs focused on physics concepts. Examples of quiz items included "What is the charge of 20 electrons?", "What does it mean for the filament of a light bulb to be in equilibrium?", and "A student measures a current of 0.04 A in a nichrome wire with length  $L$ . What should the student measure in a nichrome wire of length  $5L$ ?" The TAs who taught the conceptual inquiry-based labs began their lab sessions with 5-minute introductory lectures in which they introduced physics concepts by drawing diagrams and writing equations on the whiteboard at the front of the room.

By comparison, the skills-based labs used apparatus that mirrored the conceptual inquiry-based labs, but focused on experimental skills instead of physics concepts. For example, in Lab 5, the conceptual inquiry-based lab used two different gauges of nichrome wire in order to help students develop an understanding that the thickness, length, and applied emf of a wire will determine the electric field in that wire, which will in turn determine the amount of current that flows through the wire. In the skills-based version of Lab 5, students were presented with a D-cell battery, ammeters/voltmeters, and a box of nichrome wires with different lengths and cross-sectional areas and tasked with creating a plot that showed how the resistance varied with either cross-sectional area or length. The skills-based lab, therefore, focused on controlling variables and graphing data. For their lab reports, students wrote paragraph-long responses to questions related to experimental practices (e.g., "If you wanted to determine the resistance of a resistor, what would you need to measure?" and "How did you control variables in your investigation?") and inserted graphs with error bars, best-fit lines, and linearization, as ap-

plicable.

The pre-lab and post-lab quizzes for the skills-based labs focused on scientific thinking and evaluating scientific knowledge claims using the skills learned in the lab. Examples of quiz items included "Do the data on this plot of ice cream consumption vs. PISA score [showing a weak but positive relationship] justify the claim that eating ice cream makes children smarter?", "What quantity is represented by the slope of a linear best-fit line on the graph you will make in this lab?", and "Does this graph (showing monthly average temperatures from January to June) provide evidence for global warming?" The TAs who taught the skills-based labs began their lab sessions with 5-minute introductory lectures in which they discussed the relevant experimental skills, such as creating and reading graphs, using error bars to compare measured values, and evaluating scientific knowledge claims.

These two lab curricula were assembled specifically for this study to be as similar as possible, with one significant difference: the conceptual inquiry-based lab sought to help students develop their understanding of physics concepts through PER-informed lab-work, and the skills-based lab sought to help students develop their understanding of experimental practices and ways of thinking without focusing on physics concepts. In order to determine the extent to which the conceptual inquiry-based lab helped students develop their understanding of physics concepts, students completed the BEMA at the start of their lab periods during the first and last labs of the semester. Since the students were simultaneously learning about electricity and magnetism in their lecture class, the skills-based lab was needed to serve as a control group. The BEMA includes 31 questions but, because two questions are combined when evaluating responses, returns a score from 0 to 30.

A study of students at similar institutions suggests a significant increase in BEMA score for students who are enrolled in a *Matter & Interactions* course [23]. From that study, I estimated an increase of approximately 5 points to the BEMA score for students enrolled in a *Matter & Interactions* lecture course. A power analysis [24] allowing for 5% false-positive and false-negative rates indicated that a study with 300 students would be able to rule out differences between the two lab curricula at the size of about 1 point on the 30-point BEMA score. I reasoned that a gain of less than 1 point in the one-credit lab course would provide poor value compared with a gain of 5 points in the three-credit lecture course. Therefore, since the power analysis indicated that this study would be able to provide a definitive answer to the question of whether a conceptual inquiry-based lab would have an educationally significant impact on student conceptual learning, I went ahead with the study.

A first-pass analysis of the data included descriptive and comparative statistics to look for substantive differences between students in the conceptual inquiry-based labs and the skills-based labs. However, due to the nested nature of the instruction (i.e., students had different TAs, who were randomly assigned to one of the two lab curricula), a hierarchical linear

TABLE I. Descriptive and comparative statistics for this study. Column 2 indicates the number of study participants, columns 3-5 indicate BEMA scores (from 0 to 30) and their change, with standard errors. Comparative statistics in this table include Hake's  $g$ , Cohen's  $d$ , and the  $p$ -value associated with Student's t-test [25]. In both types of labs, student scores increased significantly, and by approximately the same amount. There were significant differences between the Pre and Post scores for students in the two different lab curricula (rows 4-5).

	N	Pre (SE)	Post (SE)	Change (SE)	$g$	$d$	$p$
conceptual inquiry-based labs	159	9.31 (0.23)	14.54 (0.38)	5.22 (0.37)	0.25	1.31	< 0.001
skills-based labs	104	10.25 (0.34)	15.50 (0.54)	5.25 (0.43)	0.27	1.13	< 0.001
$d$		-0.30	-0.19	-0.005			
$p$		0.012	0.075	0.483			

model (HLM) [26] is necessary to adequately account for potential instructor-level effects. A linear model also permits the use of a difference-in-differences strategy [27], which allows the extraction of a treatment effect from data that might include parallel trends.

Since difference-in-differences has not yet been used widely in PER, I will briefly outline the concept. Imagine that a sample of participants is assigned to either the treatment or control condition. Over time, there may be a change to the outcome variable in addition to any change that might arise because of the treatment. I assume that while the outcome variable may initially be different between the treatment and control groups, the evolution of the outcome variable is such that the change in the value for the control group is the same as the change would have been for the treatment group if the treatment group had not received treatment. In such case, I can estimate the impact of the treatment by calculating

$$Effect = (Post - Pre)_{Treatment} - (Post - Pre)_{Control}$$

For example, if students attending an extra help session saw their FCI scores jump from 14 to 19, on average, while students in the regular class saw a jump from 15 to 18, I might conclude that the extra help session caused FCI scores to increase by  $(19 - 14) - (18 - 15) = 2$  points.

In order to apply the difference-in-differences technique to a hierarchical linear model, I start with the model definition. The BEMA score ( $Score_{ij}$ ) of student  $i$ , in a lab section taught by TA  $j$ , will be predicted by the fixed effects of time ( $Time_{ij} = 0$  for the pre, and 1 for the post) and the curriculum ( $Curriculum_j = 0$  for skills-based and 1 for conceptual inquiry-based labs).

#### Level 1

$$Score_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + r_{ij}$$

#### Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Curriculum_j + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Curriculum_j + U_{1j}$$

The  $\gamma$  coefficients may thus be interpreted in terms of the difference-in-differences framework, controlling for random effect and random slopes associated with the different TAs.  $\gamma_{00}$  is the average BEMA pre score for students in

the skills-based lab.  $\gamma_{01}$  is the difference in average BEMA pre scores between the conceptual inquiry-based and skills-based labs.  $\gamma_{10}$  is the average pre-to-post increase in BEMA score for students in the skills-based labs. And, finally,  $\gamma_{11}$  is the difference-in-differences effect estimation, the average increase in BEMA post scores for students in conceptual inquiry-based labs after subtracting off the pre-to-post difference in the skills-based labs. The scores have not been normalized, so  $\gamma_{11}$  can be read immediately as the average increase in number of points on the BEMA score that can be expected for students in the conceptual inquiry-based labs compared with the skills-based labs.

Lastly, I want to briefly describe the ethical decisions that were made in the course of this research project, using the framework laid out in the Belmont Report [28]. Considering the need to respect the autonomy of our research participants, I decided not to collect personal, background, or demographic data from participants as I could not justify the need for that data to answer the research question. Considering the need for beneficence, I worked to develop a lab course for students that was more engaging and educationally valuable than they would otherwise have encountered. I resisted the urge to use a 'business as usual' lab course as the control group, and made sure all students – whether they decided to opt-in to the study or not – got a better education than they would have received if I had not run this study. Considering the need to respect persons, I declined to include any additional surveys or questions than would be absolutely necessary to answer the research question. I integrated the post BEMA with the end-of-semester survey and TA evaluations in order to minimize the number of surveys students were being asked to complete. Finally, while I acknowledge that institutional review marks a minimum rather than aspirational standard, I note that this study was approved by the Institutional Review Board at North Carolina State University.

### III. RESULTS

Two analyses were conducted with the data. First, descriptive and comparative statistics were calculated to compare BEMA score gains for the two curricula. These statistics, presented in Table I, indicate that the two types of labs saw

significant and comparable gains in BEMA scores. The average BEMA score in conceptual inquiry-based labs increased by 5.22 points (SE = 0.37, Hake’s [29]  $g = 0.25$ , Cohen’s [30]  $d = 1.31$ ) while the average BEMA score in skills-based labs increased by 5.25 points (SE = 0.43,  $g = 0.27$ ,  $d = 1.13$ ).

Second, the difference-in-differences analysis using HLM was evaluated using the nlme package in R [31]. Results are presented in Table II. Matching the approximate gains reported previously for students in Matter & Interactions courses [23], I find an average pre-to-post gain of 5.25 (SE = 0.59) points in BEMA scores for students in the skills-based labs. The average difference in BEMA pre scores between the two lab curricula was 0.94 (SE = 0.55) points lower for students in the conceptual inquiry-based labs, which is not statistically significant at  $p < 0.05$ .

The primary result is that the difference-in-differences estimator  $\gamma_{11}$  is not statistically different from zero, corresponding to a difference of 0.02 (SE = 0.76) points. Therefore, I accept the null hypothesis that there is no added benefit to the conceptual inquiry-based labs, compared with the skills-based labs.

TABLE II. HLM results for this study. The second column indicates model estimates for the intercept and fixed effects described in the text. The difference-in-differences effect estimate,  $\gamma_{11}$ , is not statistically different from zero, indicating that there was no difference in average BEMA gains between students in the skills-based and conceptual inquiry-based labs.

Fixed Effect	Coefficient	SE	$p$
Intercept ( $\gamma_{00}$ )	10.25	0.42	< 0.001
Curriculum ( $\gamma_{01}$ )	-0.94	0.55	0.156
Time ( $\gamma_{10}$ )	5.25	0.59	< 0.001
Curriculum x Time ( $\gamma_{11}$ )	-0.02	0.76	0.975

#### IV. DISCUSSION AND CONCLUSION

Recent scholarship has made a strong case that traditional physics labs do little to support student learning of physics concepts [8, 9]. The result of the present study, that BEMA scores increased no more for students in conceptual inquiry-based labs than for students in skills-based labs, seems to be in agreement. However, I caution against extrapolating these results to claim that no introductory physics lab curriculum is capable of substantially helping students to learn physics concepts, especially in light of studies that indicate gains on concept inventory scores for students in certain types of conceptual inquiry-based lab. Instead, it may be useful to reflect on which aspects of the conceptual inquiry-based lab in this study may have proven ineffective at helping students learn physics concepts, especially in comparison with previous studies that have shown positive results.

The conceptual inquiry-based labs used in this study did not focus on multiple representations [32]. By contrast, the RealTime Physics labs, which have been associated with a

significant increase in student conceptual understanding [20], employ multiple representations frequently [15]. Use of multiple representations may help students make connections between abstract concepts (e.g., electric potential difference), diagrams (e.g., circuit diagrams), and the physical objects and measurements they will encounter in the lab. Similarly, one study that reported an increase in student conceptual understanding in an inquiry-based lab course described integrating interactive simulations [16], one more form of representation.

Another aspect of that successful inquiry-based lab course was “elicit[ing] student conceptions whenever possible” [16]. This strategy is also used by RealTime Physics, which frequently calls on students to make predictions prior to conducting different measurements [15]. Research using clickers in a lecture classroom has shown that asking students to commit to a prediction before they are exposed to a correct explanation improves their ability to learn new concepts [33]. Eliciting predictions may be an effective pedagogical strategy in labs that focus on conceptual learning.

In this study, the social environments in the conceptual inquiry-based labs and the skills-based labs were not different. In both labs, students worked with a partner to assemble a lab report, and completed individual pre-lab and post-lab quizzes. However, in two other studies, the conceptual inquiry-based labs (but not the traditional labs) were re-worked to enhance the value of collaborative learning, either by explicitly prompting discussion between lab partners [17] or by requiring that students meet and design an experimental procedure prior to the lab [18]. It is possible that the effectiveness of conceptual inquiry-based labs can be unlocked if productive collaboration and cooperation are encouraged, scaffolded, and supported.

There are several limitations to this study. There is a wide variety of conceptual inquiry-based lab curricula, and this study focused on only one such curriculum. Likewise, large universities like the one at which this study was performed are over-represented in PER [34], and it is not necessarily clear that these results would be the same at a two-year college or a high school, for example. The lab studied here was focused on electricity and magnetism, while much prior scholarship has centered on mechanics-focused labs, which may limit the usefulness of comparisons with other scholarship.

This randomized controlled trial compared BEMA pre-to-post score gains for students assigned to either a conceptual inquiry-based lab or a skills-based lab in an introductory physics lab course at a large research-intensive university in the USA South. There was no difference in average pre-to-post gain between the students in the two labs.

#### ACKNOWLEDGMENTS

I thank the six excellent graduate student TAs (Aidan, Anjali, Dip, Gregory, Jacob, and Theo) who brought these two lab curricula to life. I thank the labs committee at NCSU for support and advice. I acknowledge the labor and friendship of Marceline and Claudia for managing our facilities.

- 
- [1] A. A. Bless, Cook-book laboratory work, *The American Physics Teacher* **1**, 88 (1933).
- [2] R. T. White, Relevance of practical work to comprehension of physics, *Physics Education* **14**, 384 (1979).
- [3] W. S. Toothacker, A critical look at introductory laboratory instruction, *American Journal of Physics* **51**, 516 (1983).
- [4] K. Wagoner, K. M. Hynes, and D. Flanagan, Interesting guided-inquiry labs for a large-enrollment, active learning physics ii course, *The Physics Teacher* **56**, 244 (2018).
- [5] H. Kruglak, Achievement of physics students with and without laboratory work, *American Journal of Physics* **21**, 14 (1953).
- [6] D. D. Long, G. W. McLaughlin, and A. M. Bloom, The influence of physics laboratories on student performance in a lecture course, *American Journal of Physics* **54**, 122 (1986).
- [7] C. Wieman and N. Holmes, Measuring the impact of an instructional laboratory on the learning of introductory physics, *American Journal of Physics* **83**, 972 (2015).
- [8] N. G. Holmes, J. Olsen, J. L. Thomas, and C. E. Wieman, Value added or misattributed? A multi-institution study on the educational benefit of labs for reinforcing physics content, *Phys. Rev. PER* **13**, 010129 (2017).
- [9] E. M. Smith, M. M. Stein, C. Walsh, and N. G. Holmes, Direct measurement of the impact of teaching experimentation in physics labs, *Phys. Rev. X* **10**, 011029 (2020).
- [10] N. G. Holmes and H. J. Lewandowski, Investigating the landscape of physics laboratory instruction across north america, *Phys. Rev. Phys. Educ. Res.* **16**, 020162 (2020).
- [11] R. Chambers, What use are practical physics classes?, *Phys. Bull.* **14**, 181 (1963).
- [12] A. B. Arons, Guiding insight and inquiry in the introductory physics laboratory, *The Physics Teacher* **31**, 278 (1993), <https://doi.org/10.1119/1.2343763>.
- [13] I thank historian Joanna Behrman for pointing out that a description of lab-work as ‘cookbook’ draws on a sexist view of scientific work [35], and also thank Dimitri Dounas-Frazer for his advocacy on this matter.
- [14] F. Read, New techniques for the teaching of practical physics, *Physics Education* **4**, 77 (1969).
- [15] D. R. Sokoloff, P. W. Laws, and R. K. Thornton, RealTime Physics: Active learning labs transforming the introductory laboratory, *Eur. J. Phys.* **28**, S83 (2007).
- [16] D. J. Van Domelen and A. Van Heuvelen, The effects of a concept-construction lab course on FCI performance, *American Journal of Physics* **70**, 779 (2002).
- [17] T. Chambers, *Three pedagogical approaches to introductory physics labs and their effects on student learning outcomes*, Ph.D. thesis, University of Arizona (2014).
- [18] M. Riaz, T. J. Marcinkowski, and A. Faisal, The effects of a DLSCL approach on students conceptual understanding in an undergraduate introductory physics lab, *Eurasia Journal of Mathematics, Science and Technology Education* **16** (2020).
- [19] G. A. B. Nock, *The effects on community college student physics achievement and attitudes about learning physics due to inquiry-based laboratory activities versus cookbook laboratory activities*, Ph.D. thesis, University of Mississippi (2009).
- [20] D. Doucette, R. Clark, and C. Singh, Students’ attitudes toward experimental physics in a conceptual inquiry-based introductory physics lab, *Canadian Journal of Physics* **10.1139/cjp-2021-0121** (2022), Just-IN.
- [21] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010105 (2006).
- [22] R. W. Chabay and B. A. Sherwood, *Matter and Interactions*, 4th ed. (John Wiley & Sons, 2015).
- [23] M. A. Kohlmyer, M. D. Caballero, R. Catrambone, R. W. Chabay, L. Ding, M. P. Haugan, M. J. Marr, B. A. Sherwood, and M. F. Schatz, Tale of two curricula: The performance of 2000 students in introductory electromagnetism, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020105 (2009).
- [24] ClinCalc, Sample size calculator, <https://clincalc.com/stats/samplesize.aspx>.
- [25] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures* (CRC Press, 2003).
- [26] B. Van Dusen and J. Nissen, Modernizing use of regression models in physics education research: A review of hierarchical linear modeling, *Phys. Rev. Phys. Educ. Res.* **15**, 020108 (2019).
- [27] J. D. Angrist and J.-S. Pischke, *Mastering Metrics: The path from cause to effect* (Princeton University Press, 2014).
- [28] N. C. for the Protection of Human Subjects of Biomedical and B. Research, *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (Department of Health, Education, and Welfare, 1978).
- [29] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *American Journal of Physics* **66**, 64 (1998).
- [30] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 2013).
- [31] J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS* (Springer, New York, 2000).
- [32] N. Munfaridah, L. Avraamidou, and M. Goedhart, The use of multiple representations in undergraduate physics education: What do we know and where do we go from here?, *Eurasia Journal of Mathematics, Science and Technology Education* **17**, em1934 (2021).
- [33] C. Crouch, A. P. Fagen, J. P. Callan, and E. Mazur, Classroom demonstrations: Learning tools or entertainment?, *American Journal of Physics* **72**, 835 (2004).
- [34] S. Kanim and X. C. Cid, Demographics of physics education research, *Phys. Rev. Phys. Educ. Res.* **16**, 020106 (2020).
- [35] J. Behrman, The Origin of ‘Cookbook’ Laboratories (2017), AAPT Winter Meeting.