Tools for identifying courses that support development of expertlike physics attitudes

Jayson M. Nissen,[1] Ian Her Many Horses,[2] Ben Van Dusen,[3]
Manher Jariwala,[4] and Eleanor W. Close[5]

[1]*Nissen Education Research and Design, Corvallis, Oregon 97333, USA*
[2]*School of Education, University of Colorado Boulder, Boulder, Colorado 80309, USA*
[3]*School of Education, Iowa State University, Ames, Iowa 50011, USA*
[4]*Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[5]*Department of Physics, Texas State University, San Marcos, Texas 78666, USA*

(Received 3 December 2020; accepted 10 March 2021; published 13 April 2021)

Educators and researchers often use research-based assessments before and after instruction to measure the efficacy of courses. Limited resources exist for interpreting assessment results, particularly for attitudinal surveys. We present analyses and representations created with data from 192 introductory physics courses that educators and researchers can use to provide a context for interpreting results from the Colorado Learning Attitudes About Science Survey. The provided data came from the online Learning About STEM Student Outcomes platform and from the scientific literature. The representations include scatter plots of pretest and post-test scores and distributions of effect sizes. Educators and researchers can use these representations to show how courses compare to the larger database before and after instruction. Results almost always associated physics courses for nonscience majors with shifts to more expertlike attitudes, whereas this was the case for approximately one-third of introductory physics courses for science majors.

DOI: 10.1103/PhysRevPhysEducRes.17.013103

I. INTRODUCTION AND MOTIVATION

Professional physicists tend to hold a set of attitudes about what it means to learn and do physics [1–3]. Many of these cultural attitudes in physics, such as valuing competition, individualism, solitary practice, and genius, have greater costs for women and Black, Indigenous, and people of color [4–6]. In this article, we focus on cultural attitudes beneficial for learning and integration into the physics community [7,8] such as viewing physics as applicable to daily life, perceiving personal effort as supporting learning physics, and taking approaches to solving and understanding physics problems that value sense making over memorization [9]. Students who do not start college holding a vast majority of these attitudes seldom become physicists in the current educational system [1,2]. Yet, physics instructors often rank attitudes toward and appreciation of physics as less important course outcomes than conceptual understanding and problem solving [10]. That lack of value likely plays a role in how many college physics courses, particularly courses for physical science majors, and degree programs do not support attitude development [1,2,11]. These characteristics of physics education imply that physics programs filter out students who do not already hold these attitudes when they enter college physics courses [1,2]. This filtering effect disproportionately harms women and Black, Indigenous, and people of color [12]. Supporting students in developing expertlike attitudes may support them in having a more enjoyable and productive experience in their physics courses and may support physics departments in recruiting and retaining more and more diverse physics majors.

Research-based curriculum focused on conceptual learning are not sufficient for supporting students in developing beneficial attitudes about physics [13]. Instead, curriculum that improve these attitudes such as Physics and Everyday Thinking for future educators [14] and Modeling Instruction in courses for science majors [15,16] explicitly focus on attitude development and conceptual learning as mutual goals. The Physics by Inquiry curriculum, which also improves student attitudes, implicitly focuses on attitude development while explicitly emphasizing conceptual learning [17]. These results show that students' attitudes become more expertlike in courses that emphasize attitudinal development and conceptual learning. Most algebra- and calculus-based physics courses, however, do not explicitly target attitudes about science [10].

To measure shifts in students' attitudes about physics, educators can administer the Colorado Learning Attitudes about Science Survey (CLASS) [3] prior to and after instruction. The CLASS examines attitudes about the nature of physics knowledge and its relation to everyday

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

life, ideas about learning and problem solving, and attitudes toward sense making [9,18]. Instructors can compare students' responses to those given by physics experts to indicate the extent to which students' attitudes become more expertlike following instruction. Measuring students developing expertlike attitudes can inform pedagogical choices, help convince colleagues to adapt or improve their teaching practices, persuade administrators to sustain or expand course transformations, and guide the development of evidence-based instructional strategies.

Many research-based assessments like the CLASS have normative data available in the literature that provide context for interpreting local results. Although some publications provide comparison data for the CLASS [11], the literature overrepresents research-intensive, selective universities with less diverse and better mathematically prepared students [19]. Therefore, these studies may not generalize to more diverse student populations and institutional contexts. To address these limitations in publications, we used data (see Supplemental Material [20]) from both the literature and the online Learning About STEM Student Outcomes (LASSO) platforms. For the CLASS, many researchers only report the shifts in scores and not the scores at the beginning and end of instruction [11]. Not reporting scores from the beginning of instruction may hide biases in the literature (i.e., the scores may only come from selective, research-intensive institutions and calculus-based courses [19]). This reporting practice means that educators cannot situate their courses in terms of students' initial attitudes within much of the existing literature.

This article provides a large-scale dataset of CLASS scores and tools for interpreting the impact of courses on students' self-reported attitudes about physics as compared to experts' attitudes in the Supplemental Material [20]. The dataset includes courses from the LASSO platform and from research literature reporting sufficiently detailed student outcome data. The tools include a scatter plot of pretest and post-test CLASS scores for all physics courses in the dataset and a distribution of the effect sizes for the shifts on students' attitudes for all courses. The Supplemental Material includes an Excel file and an R file with the data and code for readers to create their own visualizations to understand and communicate their courses impacts on students' attitudes [20].

II. PURPOSE

By providing benchmarks for students attitudes about physics, we hope to support educators in creating and using pedagogical practices in their classrooms that foster expertlike attitudes. These benchmarks can help identify the efficacy of innovations and provide context for sharing these innovations. Evidence-based pedagogical practices that support students in developing expertlike attitudes benefit students, educators, and departments. By improving students' experiences and learning outcomes in physics courses, these curriculum support departments in recruiting and supporting more, more diverse, and better prepared students.

III. METHODS

A. Data collection

LASSO is an online assessment platform hosted on the Learning Assistant Alliance website [21] that collects two levels of data: course data and student data. Instructors describe the course context, teaching method, and provide a student roster with emails when setting up the assessment. LASSO then allows instructors to email assessment invitations to students to take the assessment online. Information collected from the students includes demographic data, assessment metadata (e.g., time taken on the assessment), and responses to the individual questions. LASSO provides instructors with a report that includes summary statistics and visualizations of student performance. The platform anonymizes the data and makes it available to researchers only for students who agree to share their data for research.

We used Google Scholar [22] to search all articles that cited the initial CLASS publication (Adams *et al.* [3]) and a review of attitudes in physics (Madsen *et al.* [11]). This search provided 26 articles or dissertations that reported pretest and post-test mean scores. We included the data for individual courses whenever possible. Several studies reported statistics that combined courses, which we included as a single course. The literature (48 courses with 8630 students) and LASSO data (144 courses with 7094 students) did not overlap.

We identified courses as being either calculus-based (for physical science and engineering majors) or algebra-based (for other science majors), for nonscience majors, or for future educators. We did not differentiate between introductory mechanics and electricity and magnetism courses because preliminary analyses indicated that CLASS scores were similar in these course types. Not distinguishing between mechanics and electricity and magnetism simplified the figures and tables. We excluded modern physics courses from Ref. [23] to simplify the analyses because all of the courses came from the same institution. Course descriptions and course numbers in the LASSO database along with college catalogs provided information for identifying the courses. The physics courses for future educators were predominantly for future elementary educators and most were taught using Physics by Inquiry or Physics and Everyday Thinking, which are designed to support the development of expertlike attitudes about physics.

We combined the two datasets and did not focus on differences between the literature and LASSO data. Identifying differences between the two data sources was limited by the literature only containing enough courses to

make quantitative comparisons for calculus-based physics and physics for education majors. Consistent with Kanim and Cid [19], the scores reported in the literature tended to be higher than the scores reported on LASSO: approximately 1 percentage point for calculus-based courses and 8 percentage points for courses for education majors. Comparing the shifts from pre to post as measured by effect sizes in these two course types, we found similarities between the data from LASSO and the literature: $d_{\mathrm{LASSO}} = -0.11$ and $d_{\mathrm{lit}} = -0.07$ for calculus-based courses and $d_{\mathrm{LASSO}} = 0.41$ and $d_{lit} = 0.49$ for courses for education majors. We describe calculating effect sizes below in Sec. III C. Combining the two data sources provided a larger and more diverse dataset.

We scored the CLASS responses using the 36 item agree scoring scheme recommended by the authors [3]. While studies have problematized the use of all 36 items [9] and the category collapsing used in the scoring scheme [24], we used the author's scoring recommendations as they likely align with educator's scoring practices and are used in the literature.

B. Multiple imputation

Physics education research studies often handle missing data by removing every student that does not have both a pretest and a post-test, which is often called complete case analysis. Statisticians recommend against using complete case analysis because it often biases the results and instead recommend multiple imputation or other principled methods for handling missing data [25]. Multiple imputation uses all available data by imputing the missing values multiple times to create multiple complete datasets. Multiple imputation then combines the results of the analyses of each imputed dataset to produce unbiased estimates with accurate uncertainties. Nissen *et al.* [26] showed that multiple imputation produces more accurate results than complete case analysis with RBA data for physics education research studies. Schafer [25] provides a detailed overview of multiple imputation.

To remove spurious data, we filtered the LASSO data for students that completed the assessment in under 3 min or failed to correctly answer the filter question [3]. We then used hierarchical multiple imputation [27] to impute scores for students who completed only the pretest or post-test and created a single dataset of course average scores. Averaging multiple imputed datasets is not the best practice as it creates artificially small error bars; however, it sufficed for this analysis as we did not need to propagate the errors and it allowed us to create visualizations to help educators and researchers contextualize course outcomes. Several studies in the literature only provided partial information (e.g., post-test scores or no standard deviations) and we used multiple imputation to impute their missing values after combining the course-level data from LASSO and the literature.

C. Effect size

We reported the course-level shifts from pretest to post-test using Cohen's d with Hedges correction as a measure of effect sizes, given in Eq. (1). Cohen's d calculates the shift in the mean score from pretest to post-test in standard deviations. Hedges correction addresses bias toward larger effect sizes for smaller sample sizes. Kraft [28] provides a guide to using and interpreting effect sizes.

$$d = \frac{x_{\mathrm{post}} - x_{\mathrm{pre}}}{(sd_{\mathrm{post}} + sd_{\mathrm{pre}})/2} \frac{(N-3)}{N-2.25} \sqrt{\frac{N-2}{N}}. \qquad (1)$$

IV. FINDINGS

We present the findings for CLASS scores across four types of introductory physics courses: calculus-based courses, algebra-based courses, courses for future educators, and courses for nonscience majors. The first subsection below focuses on mean pretest and post-test scores for courses. The second subsection focuses on effect sizes for courses.

A. CLASS scores across course types

Figure 1 plots the course pretest and post-test scores and distributions of course scores for the four course types.

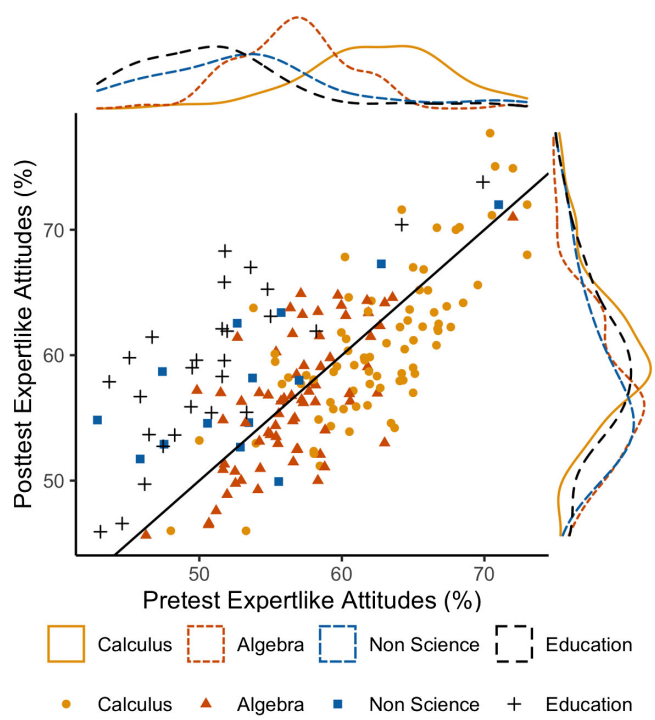

FIG. 1. Scatter and density plots for pretest and post-test attitudes in the four types of courses. Attitudes tended to start lower in courses for education and nonscience majors than in calculus- and algebra-based courses. Attitudes, however, became more expertlike in courses for education and nonscience majors to the extent that they largely overlapped with those for calculus- and algebra-based courses on the post-test. The diagonal line represents no shift from pretest to post-test.

Over all course types, 94 of 192 courses had increases in expertlike attitudes and 98 had decreases in attitudes. Expertlike attitudes increased more often in physics courses for future educators (28/28) and nonscience majors (12/14). But, attitudes also increased in approximately a third of algebra- (29/71) and calculus-based (25/79) courses. The density plots in Fig. 1 show the cumulative effect of this difference in effect size across course types. On the pretest, the density plots only slightly overlap for calculus-based physics courses and courses for either education or nonscience majors. On the post-test, however, these density plots largely overlap. This shift in distributions indicates that pedagogies, such as those focused on conceptual learning and developing expertlike attitudes in the courses for future educators and nonscience majors can have meaningful impacts on students' attitudes.

Pretest attitudes were much lower in the courses for future educators and nonscience majors where attitudes tended to become more expertlike. This relationship leads to the possibility that lower pretest CLASS scores differentiates between courses associated with either increased or decreased expertlike attitudes. The data, however, contradict this possibility. Many of the algebra- and calculus-based courses in the top half of the pretest scores for those course types had positive effect sizes: 10 of 25 for calculus-based courses and 15 of 29 for algebra-based courses. Furthermore, four courses for either education or nonscience majors had pretest means above 60% and all four had positive effect sizes.

B. Effect sizes across course types

The effect size varied across all course types ranging from a low of -0.6 to a high of 1.3 with an average of 0.04, see the distribution for all courses in the lowest density plot in Fig. 2. The averages and distributions of effect sizes, however, varied greatly across the four course types; see Table I and Fig. 2. Effect sizes were more frequently near zero or negative in calculus- and algebra-based courses and positive and often quite large in courses for education and nonscience majors. These results showed that meaningful positive effect sizes occurred in all course types and provide educators and researchers with another metric for understanding where their courses stand in comparison to both what is typical and what is possible.

V. LIMITATIONS

This synthesis was limited to courses and studies that used the CLASS. Other attitudinal and affective measurements exist in Physics Education Research. For example, Madsen *et al.* [11] provides a review of studies that used both the CLASS and the Maryland Physics Expectations Survey and Henderson *et al.* [29] provides a synthesis of articles looking at physics self-efficacy.

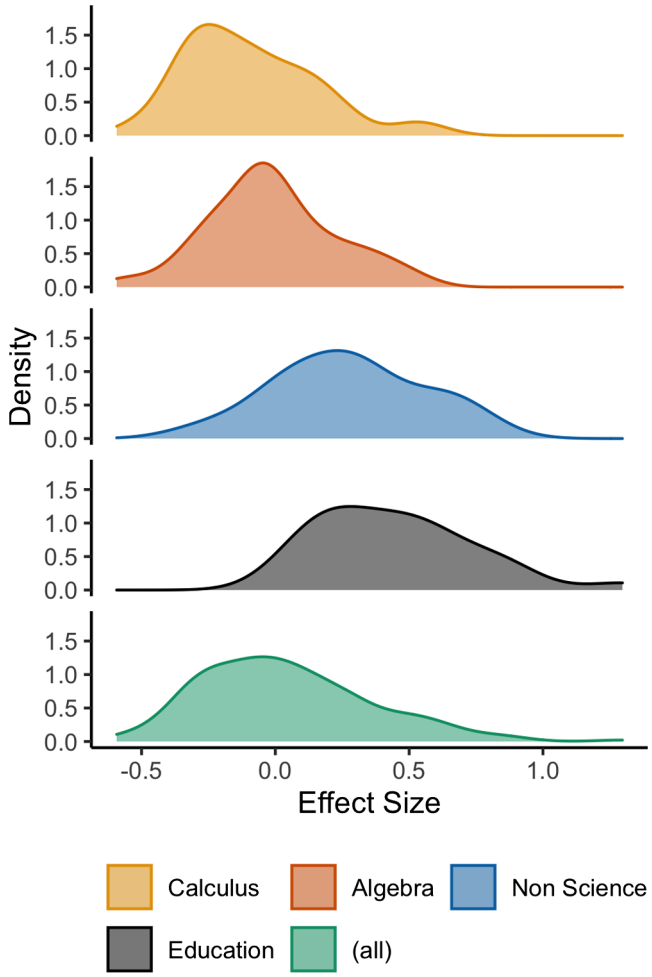

FIG. 2. Density plots showing the distributions of effect sizes for all four physics course types (calculus- and algebra-based courses and courses for education and nonscience majors) and the overall distribution of effect sizes. Approximately one-third of algebra- and calculus-based courses had shifts toward more expertlike attitudes while nearly all courses for education and nonscience majors did.

The CLASS also presents several limitations. Several of the authors (J. N., I. H. M. H., and B. V. D.) discuss the CLASS in detail in Ref. [12] and we do not restate those arguments here because many are outside the scope of this

TABLE I. Descriptive statistics for the four types of introductory physics courses: calculus-based courses, algebra-based courses, courses for future educators, and courses for nonscience majors.

		Effect size		Pretest (%)		Post-test (%)	
Course	N	Mean	S. D.	Mean	S. D.	Mean	S. D.
Calculus	79	−0.09	0.25	62.54	15.57	60.79	16.67
Algebra	71	−0.01	0.24	56.71	15.43	56.42	16.08
Non Sci.	14	0.27	0.28	53.50	15.95	57.95	16.43
Education	28	0.45	0.30	51.15	17.81	59.19	18.12

article. The original development of the CLASS focused on attitudes physicists found important and was only loosely based on theoretical frameworks [9]. Subsequent work has identified a three-factor model as the best fit for the CLASS [9,18]. However, even these articles place little emphasis on the theoretical foundations for these three latent variables. For example, Kontro and Buschhüter [18] labels one as problem solving self-efficacy but makes no reference to Bandura *et al.* [30] or other relevant literature on self-efficacy. Grounding measurements in theoretical frameworks guides interpreting those measurements meaning and can inform the design and implementation of pedagogies and interventions that support students developing expertlike attitudes. For example, a measurement construct built to align with self-efficacy [30] and having validity evidence for measuring self-efficacy would allow researchers to draw on the broader social cognitive theory [31,32] that contains self-efficacy to design, implement, and test interventions with confidence in the alignment between the measurement tool and the intervention. Theoretical foundations align interventions and measurements to focus efforts on developing beneficial interventions and pedagogies. Even though Douglas *et al.* [9] and Kontro and Buschhüter [18] disagree with Adams *et al.* [3] about the number of factors and number of items to use on the CLASS, we are confident that the overall score used in this study would correlate with those subconstructs and provides insight on the general change in attitudes in college physics courses.

VI. DISCUSSION

By describing 192 courses from the literature or LASSO database, these results show the state of attitudinal outcomes across four common types of physics courses. This synthesis shows both what is possible, shifts towards expertlike attitudes in almost every course for future educators and nonscience majors, and what can be improved, courses for science majors. This synthesis also acts as a benchmark to better understand course outcomes. For educators and researchers who develop novel methods for supporting students in gaining more expertlike attitudes, these results can provide a context for motivating their work and interpreting that work's efficacy.

The results demonstrate a stark difference between the shifts in students' attitudes for introductory physics courses for future educators and non-science majors compared to algebra- and calculus-based courses. The attitude development in courses for future educators and nonscience majors indicate that focusing on attitudes can support students' developing expertlike attitudes. The beneficial relationship between attitudes and conceptual learning, becoming a physics major, and the quality of student's experiences in physics all warrant a greater emphasis on attitudes in algebra- and calculus-based physics courses. Physics courses for science majors have the potential to improve a broad swath of student outcomes, including the experiences that students have in their physics courses, by developing and using evidence-based pedagogies that support students' attitudes becoming more expertlike. This lack of focus on attitudes likely plays a role in why so many highly competent students leave physics and other science majors in pursuit of more meaningful degrees [5]. Courses that emphasize attitude development can support departments in attracting and retaining more and more diverse students. For the students who take these courses and decide not to become science majors, they can still develop a deeper appreciation of the role of science in society and everyday life.

ACKNOWLEDGMENTS

The work was supported by NSF DUE 1928596 and 1525354.

[1] E. Gire, B. Jones, and E. Price, Characterizing the epistemological development of physics majors, Phys. Rev. ST Phys. Educ. Res. **5**, 010103 (2009).

[2] S. P. Bates, R. K. Galloway, C. Loptson, and K. A. Slaughter, How attitudes and beliefs about physics change from high school to faculty, Phys. Rev. ST Phys. Educ. Res. **7**, 020114 (2011).

[3] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. ST Phys. Educ. Res. **2**, 010101 (2006).

[4] M. Ong, Body projects of young women of color in physics: Intersections of gender, race, and science, Soc. Probl. **52**, 593 (2005).

[5] E. Seymour and Nancy M. Hewitt, *Talking about Leaving* (Westview Press, Boulder, CO, 1997).

[6] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).

[7] L. Ding, Verification of causal influences of reasoning skills and epistemology on physics conceptual learning, Phys. Rev. ST Phys. Educ. Res. **10**, 023101 (2014).

[8] K. K. Perkins, W. K. Adams, S. J. Pollock, N. D. Finkelstein, and C. E. Wieman, Correlating student beliefs with student

learning using the Colorado Learning Attitudes About Science Survey, AIP Conf. Proc. **790**, 61 (2005).

[9] K. A. Douglas, M. S. Yale, D. E. Bennett, M. P. Haugan, and L. A. Bryan, Evaluation of Colorado Learning Attitudes About Science Survey, Phys. Rev. ST Phys. Educ. Res. **10**, 020128 (2014).

[10] C. Henderson and M. H. Dancy, Impact of physics education research on the teaching of introductory quantitative physics in the United States, Phys. Rev. ST Phys. Educ. Res. **5**, 020107 (2009).

[11] A. Madsen, S. B. McKagan, and E. C. Sayre, How physics instruction impacts students beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).

[12] J. Nissen, I. Her Many Horses, and B. Van Dusen, Investigating society's educational debts due to racism and sexism in student attitudes about physics using quantitative critical race theory, Phys. Rev. Phys. Educ. Res. **17**, 010116 (2021).

[13] A. Elby, Helping physics students learn how to learn, Am. J. Phys. **69**, S54 (2001).

[14] V. K. Otero and K. E. Gray, Attitudinal gains across multiple universities using the physics and everyday thinking curriculum, Phys. Rev. ST Phys. Educ. Res. **4**, 020104 (2008).

[15] E. Brewe, L. Kramer, and G. O'Brien, Modeling instruction: Positive attitudinal shifts in introductory physics measured with class, Phys. Rev. ST Phys. Educ. Res. **5**, 013102 (2009).

[16] A. Traxler and E. Brewe, Equity investigation of attitudinal shifts in introductory physics, Phys. Rev. ST Phys. Educ. Res. **11**, 020132 (2015).

[17] B. A. Lindsey, L. Hsu, H. Sadaghiani, J. W. Taylor, and K. Cummings, Positive attitudinal shifts with the physics by inquiry curriculum across multiple implementations, Phys. Rev. ST Phys. Educ. Res. **8**, 010102 (2012).

[18] I. Kontro and D. Buschhüter, Validity of colorado learning attitudes about science survey for a high-achieving, finnish population, Phys. Rev. Phys. Educ. Res. **16**, 020104 (2020).

[19] S. Kanim and X. C. Cid, Demographics of physics education research, Phys. Rev. Phys. Educ. Res. **16**, 020106 (2020).

[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.17.013103 for an Excel spreadsheet with all of the data and sheets for generating several of the figures and an R Markdown document with code to recreate all of the figures.

[21] Learning Assistant Alliance, Learning about stem student outcomes (LASSO) platform (2020).

[22] Google, Google Scholar (2020).

[23] S. B. McKagan, K. K. Perkins, and C. E. Wieman, Reforming a large lecture modern physics course for engineering majors using a per-based design, AIP Conf. Proc. **883**, 34 (2007).

[24] B. Van Dusen and J. M. Nissen, Criteria for collapsing rating scale responses: A case study of the CLASS, in *Proceedings of the 2019 Physics Education Research Conference* (AIP, New York, 2019).

[25] J. L. Schafer, Multiple imputation: A primer, Stat. Methods Med. Res. **8**, 3 (1999).

[26] J. Nissen, R. Donatello, and B. Van Dusen, Missing data and bias in physics education research: A case for using multiple imputation, Phys. Rev. Phys. Educ. Res. **15**, 020106 (2019).

[27] M. Speidel, J. Drechsler, and S. Jolani, R package hmi: A convenient tool for hierarchical multiple imputation and beyond, Tech. Rep. (IAB-Discussion Paper, 2018).

[28] M. A. Kraft, Interpreting effect sizes of education interventions, Educ. Res. **49**, 241 (2020).

[29] R. Henderson, V. Sawtelle, and J. Micheal Nissen, Gender & self-efficacy: A call to physics educators, Phys. Teach. **58**, 345 (2020).

[30] A. Bandura and R. Lightsey, *Self-Efficacy: The Exercise of Control* (Freeman, New York, 1999).

[31] A. Bandura, *The Evolution of Social Cognitive Theory, Great Minds in Management* (Oxford University Press, Oxford, 2005), pp. 9–35.

[32] A. Bandura, Social cognitive theory: An agentic perspective, Asian J. Soc. Psychol. **2**, 21 (1999).