Jayson M. Nissen,[1] Ian Her Many Horses,[2]

of one study that discussed differences across race or ethnicity [14]. Traxler and Brewe [14] found gender and ethnic differences favoring men and overrepresented ethnicities (i.e., Asian and White [34]). They also found Modeling Instruction, an evidence based pedagogy with a focus on developing student attitudes, supported women and BIPOC students in developing more expertlike attitudes.

The disparities in outcomes for women and BIPOC students in physics courses result from systemic barriers in physics education. These barriers perpetuate the educational debts society owes these students [35]. Society has accrued educational debts that it owes to minoritized students through historical, sociopolitical, economic, and moral forms of inequalities [36]. In this investigation we examined an avenue by which the racist and sexist power structures within university physics courses perpetuate and increase the educational debts society owes women and BIPOC students through the denial of opportunities and resources to develop as physicists [10,37]. To better understand the role of attitudes in the lack of diversity in physics, we used a critical quantitative framework (QuantCrit) [38] to investigate the intersecting relationships between racism and sexism in inequities in student attitudes about learning and doing physics. We modeled society's educational debts due to racism, sexism, and their intersection in a multi-institutional dataset (18 institutions and 95 courses) collected using the Colorado Learning Attitudes about Science Survey (CLASS) [11] using hierarchical models. Our QuantCrit framework guided our work in an attempt to be antiracist and antisexist and counter racist and sexist uses of quantitative research in the past and present.

II. E EA CH E I

for their male peers decreased from before to after instruction in both courses. Good and colleagues' study differs from other referenced work because it focused explicitly on attitudes toward problem solving.

B.

Significant research efforts have focused on understanding the low representation of women in physics, which has remained at approximately 20% for the last 40 years [44]. Cheryan et al. [10] reviewed the literature on gender differences across the science, technology, engineering, and mathematics (STEM) domains and found that masculine cultures, gender differences in self-efficacy, and a lack of early educational experiences in the disciplines explained the lower rates of participation for women in physics, computer science, and engineering compared to biology, chemistry, and mathematics. Work in physics education research on gender differences in physics parallels Cheryan and colleagues findings. Madsen et al. [8] reviewed 26 studies on gender differences for conceptual learning in introductory physics courses. In first semester prove their skills to their peers and supervisors as well as often being viewed as “too smart” by family and friends. Clancy et al. [53] found that women of color uniquely faced barriers in astronomy and planetary sciences that White women did not face.

I . C CE A F A E

Critical race theory (CRT) began in the 1970s and 1980s as a movement among a racially diverse group of U.S. legal scholars of color to address social injustices and racial oppression [54–56]. CRT explicitly assumes racism is ingrained in our institutional structures, focuses on the narratives and counternarratives of oppressed people, and identifies the importance of interest convergence between oppressed peoples and their oppressors in creating change [57,58]. Ladson-Billings [59] provides affirmative action as a poor example of interest convergence. Affirmative action is under ongoing attack as a benefit for Black, Indigenous, and people of color and is associated with primarily benefiting Black, Indigenous, and people of color. Affirmative action in higher education, however, has primarily benefited White women [60]

We strive to clarify that our models are not measuring innate differences in students based on their race or gender, but the impacts of multidimensional oppressive power structures on students marginalized by these social constructs. One way that we reflect this in our writing is through the explicit naming of racism and sexism in interpreting our models.

(3) Data are not neutral and cannot speak for themself.—We reject the idea that data are neutral and can speak for themself. Hegemonic assumptions can shape every stage of collecting, analyzing, and interpreting data [69]. For example, the data we analyzed in this investigation came from the Learning About STEM Student Outcomes (LASSO) platform. While the LASSO platform has been

equity, diversity, and access. They summarize the equity orientations as either (i) creating new opportunities for students from historically [87] marginalized groups but not altering the status quo of what doing science in a field means or (ii) opening new possibilities for societal transformation around what it means to do science, but are less likely to impact students achievement in school directly. Philip and Azevedo [86] call on researchers to define the equity orientation their work uses.

In this article, our orientation emphasizes supporting students now over transforming what it means to do physics and who gets included in physics. We emphasize supporting students now because we feel the subset of attitudes we focus on are good outcomes in physics courses and individual instructors can enact changes in their classrooms to make these changes happen now. The scientific literature contains multiple examples of pedagogies individual instructors can use in their courses to support students in developing the attitudes they need to succeed in physics [8]. Using pedagogies that support attitude development of students from historically marginalized groups can create interest convergences because they may also improve overall recruitment and retention of physics majors in their department. Many cultural attitudes in physics (e.g., competition, individualism, and solitary practice) are more costly for women and BIPOC students to adopt [29,39]. Our focus, however, is on pedagogies that support students in seeing physics as applicable to their lives, as understanding that physics is more than plugging numbers into the right equations, and feeling capable of learning and doing physics. We pursue these goals in our own courses. By enacting these changes now, we expand the foundation for redefining what it means to learn and do physics in two ways. First, we support physics educators in reflecting on and changing their pedagogical practices. Second, we support more students from historically marginalized groups becoming physicists. These two groups of physicists may support the broader physics community in transforming what it means to learn and do physics to create a more inclusive culture in physics.

D.

Feminist theory has shown that all knowledge is marked by those who create it [88]. To be transparent about the position of the researchers in this work in relation to the power structures under investigation, we offer positionality statements [65] for each of the authors.

The following is the first author's, J. N., positionality statement. My identity as a White, cisgendered, heterosexual, nondisabled man has provided me with power and opportunities denied to others in American society. I use my experience growing up in a poor home and as a veteran of the all-male submarine service to motivate reflecting on and working to dismantle my privilege. My work on this project was shaped by the post-positivist scientific traditions I was educated in and my activist goal to pursue scientific knowledge that can help identify and dismantle policies and systems of oppression. Because of the privilege implicit in my current identities, I brought a limited perspective to this work on racism and sexism.

The following is the second author's, I. Her Many Horses, positionality statement. I identify as a Lakota (Indigenous), cisgender, heterosexual, man and was raised on the Rosebud Reservation in South Dakota. I consider myself to be educationally privileged and am a third generation college student with many family members holding terminal degrees. I hold an undergraduate degree in computer science and a Ph.D. in education. Throughout my life I am usually the only person that looks like me anywhere I go. These experiences have driven me to use my own power to address issues of equity in whatever space I find myself.

The following is the third author's, B. V D., positionality statement. I identify as a White, cisgender, heterosexual, man with a color vision deficiency. I was raised in a pair of lower-income households but I now earn an upper-middle class income. I hold an undergraduate degree in physics and a Ph.D. in education. I am an assistant professor at a Hispanic serving institution. My experiences working with marginalized students, particularly those whom I have had the honor to mentor as learning assistants [89] and as researchers, has motivated my attempts to use my position and privilege to dismantle oppressive power structures. As someone who seeks to be an ally, it is easy to overlook my own privileges. I try to broaden my perspective through feedback from those with more diverse lived experiences than my own.

. E H D

A. I

We used data collected with the Colorado Learning Attitudes about Science Survey [11] for this study. Researchers and educators commonly measure attitudes in college physics courses using either the CLASS or the Maryland Physics Expectations Survey (MPEX) [90]. The surveys ask students about several different categories of attitudes about physics, such as the relationship between learning physics and everyday life, the effort they put into learning physics, and their approach to solving physics problems. Students respond to these questions on a five-point Likert scale from strongly agree to strongly disagree. Researchers use the instruments to create an overall score and a score for each of the categories of attitudes based on how many times the students agree or strongly agree with what expert physicists reported [11,40].

The CLASS is the most commonly used measure of attitudes and attitudes about learning and doing physics [33]. However, some researchers have raised questions about what the CLASS measures and how the CLASS is

We scored the CLASS responses using the agree categories recommended by the original authors and we only analyzed the total score of the 36 items they include in their total scores [11]. We followed their original scoring recommendations so our results would be comparable to prior research using the CLASS. Adams et al. [11] recommends not including 6 of the 42 items in the total score. One excluded item is a filter question. Experts did not consistently agree on four of these items. Two items ask about the nature of science and two others ask about learning styles. The final excluded item also asks about approaches to learning but is not discussed by Adams et al. [11]. These excluded items and the extensive process Adams et al. [11] details illustrate that expert's attitudes vary.

To clean the data, we removed the pretest or post-test score if the student took less than 3 min on the assessment or incorrectly answered the filter question [11]. We removed any courses with less than 5 pretests or 5 post-tests. After cleaning the data, we used hierarchical multiple imputation (HMI) with the hmi [102] and mice [103] packages in RStudio V. 1.1.456 to impute missing data. We only imputed values for missing pretest and post-test CLASS scores, and we did not impute missing values for gender and race to respect each student's choice to not answer these questions. HMI provided a principled method for handling missing data that maximized statistical power and minimized bias while accounting for the hierarchical structure of the data [73,104–107].

The imputed dataset included 7764 students. This imputed dataset was larger than the 4673 students used in the analysis because it included students enrolled in a variety of courses: first and second semester algebra and calculus based physics courses, LA pedagogy courses, upper division physics courses, and physics courses for education majors. The rate of missing data for this dataset was 17% on the pretest and 34% on the post-test. The imputation model included a dependent variable for the post-test and accounted for the pretest score, course type, and demographic variables and nested the students within courses. The subsequent analysis only included 4673 students enrolled in first-semester algebra-based and calculus-based introductory physics courses.

C. [illegible]

To investigate student attitudes, we developed models to predict student attitudes on the pretest and post-test and in algebra-based and calculus-based first-semester physics courses separately, which are described by CLASS_{ij} in the final model. The models were 2-level hierarchical linear models with student data in the first level and course data in the second level. Using hierarchical linear models accounted for the nested nature of the data [108,109]. We ran the models and pooled the results for the imputed datasets using the mitml [110] and lme4 [111] packages in R. The hierarchical linear model parameters were fit using the penalized least squares method.

To determine what demographic variables to include in the models, we first used a rule of thumb to only investigate scores for populations with at least 20 students total [113]. This meant that we did not include variables for transgender, Hawaiian or Pacific Islander, or Native American in our models. Because removing the students with these identities could have biased the course-level results and because some students did not include a gender or race, we combined these students into two categories: gender other and race other. This meant that the final variables used in our model, which is shown above, included woman, gender other, Black, Asian, Hispanic, White, and race other. We included interactions between variables whenever a population included more than 20 students but not for the race other and gender other groups. Hispanic is often treated as an ethnicity in the United States [114]. However, 67% of Hispanic Americans consider their Hispanic identity to be a

E. I

We do not present p values. p values depend on sample size and lead to selective reporting and selective attention [80] that can ignore injustices borne by the most underrepresented and marginalized groups of students. Our analysis, instead, focused on the point estimates and standard errors produced by the models. This decision was informed by our QuantCrit perspective, which pushed us to question common statistical practices, and aligns with recommendation from the American Statistical Association in response to scientists and scientific communities misuses of p values attitudes in the calculus-based courses varied across demographic groups and ranged from -0.1 to -5

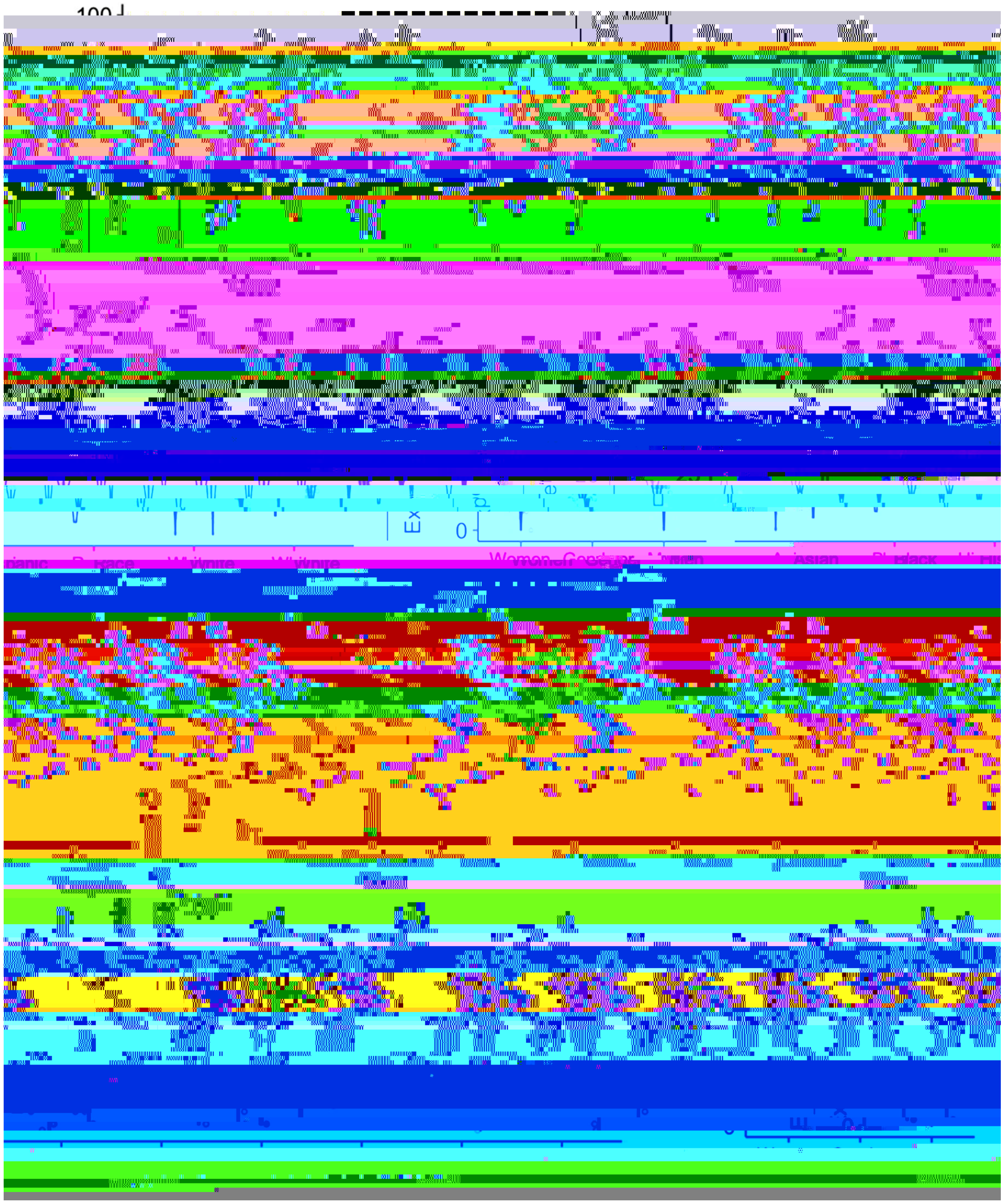
educational debt owed due to sexism across races. Figure 2 between White women and women of color were smaller than the uncertainty in the measurements. However, in both courses society owed the greatest educational debt to Black women whose predicted average attitudes were 8.1 to 12.5 percentage points lower than the predicted average attitudes for White men. To interpret the size of these differences in the context of becoming a physics major, we explore the proportion of students from each race and gender above the 75% threshold of attitudes held by most physicists in the next section.

B. 75%

As we described in the methods section, Table IV and Fig. 3 represent the results of the hierarchical generalized linear models predicting the proportion of students from each demographic group who scored above 75% on the pretest. Seventy-five percent provided an estimated cutoff for the attitudes students need prior to taking their first college physics course to have a reasonable chance of becoming a professional physicist.

The results of the hierarchical generalized linear models showed society owed educational debts to women and BIPOC students. The proportion of students making the 75% cutoff ranged from a low of 7% for Black women in algebra-based courses to a high of 36% for White men in calculus-based courses. Across all groups except Hispanic students in calculus-based courses the models predicted men to be above the threshold more often than women. In 6 of 10 comparisons, these raw differences were large. In both course types, we measured a 13 percentage point gender difference for Black students and an 8 percentage point gender difference for Asian students. In calculus-based physics courses we measured an 8 percentage point gender difference for White students and White Hispanic students. These gender differences meant Black men were more than twice as likely to be above the threshold than Black women (7% versus 20%). In the two cases where the absolute difference was smaller, for example, a 3 percentage point difference for White students in algebra-based courses, the relative difference was still large. White men were 1.2 times as likely to be above the threshold than White women in algebra-based physics courses. Most of these gender differences were much larger than the uncertainties in the measurement and the consistent gender difference across 9 of the 10 comparisons indicated society owed educational debts to women whereby men are 20% to 290% more likely to meet the threshold of more tha8.9(These)-5-1.15Mkely

convergence. The physics community should attend to



includes violin plots, box plots, and jittered scatter plots. These figures used the average values for each student for all ten imputed datasets. We included these plots because they provide data transparency for readers and because they break down the “gap gazing” perspective that means and standard errors from either descriptive statistics or statistical models can reinforce. In other words, the data show differences in the mean scores between groups (intergroup variance) but the spread of scores within each group (intragroup variance) is much larger, which is shown by

TABLE VI. Hierarchical linear model coefficients and standard errors for predicted average attitudes.

	Algebra-based				Calculus-based			
	Pre		Post		Pre		Post	
	β	SE	β	SE	β	SE	β	SE
Intercept	59.6	4.7	60.6	5.2	61.9	4.0	62.6	4.4
Gender other	−1.3	3.2	1.9	3.4	−5.5	3.0	−4.5	3.6
Hispanic	−4.0	4.4	−6.0	5.0	−3.4	3.7	−7.5	4.1
White	−0.2	4.7	−0.8	5.2	5.1	3.9	2.0	4.5
Women	−2.9	2.2	−1.8	2.6	−0.9	2.2	−1.4	2.9
Black	−4.8	5.2	−5.1	5.5	−0.5	4.1	−6.1	5.0
Asian	−2.2	4.7	−5.8	5.3	−1.6	4.0	−3.5	4.6
Race other	−3.3	4.6	−5.6	5.1	0.0	4.0	−1.6	4.7
Hisp.*White	0.3	4.4	2.1	4.9	0.9	3.7	4.1	4.4
Women*Black	−0.6	3.7	−5.0	4.4	−2.7	3.8	−2.9	4.7
Women*Asian	−0.9	2.8	0.1	3.1	−3.2	2.7	−1.6	3.4
Women*Hispanic	1.3	2.1	2.8	2.4	−1.4	2.1	0.4	2.7
Women*White	−0.2	2.2	−3.1	2.7	−2.2	2.3	−2.3	2.9

the overlap in the distributions across all groups. Across all the plots in Fig. 4, several features are worth noting: differences but not gaps across groups, medians that tend to be above 50%, and a negative skew (downward) for many of the distributions.

B.

Tables VI and VII present the model outputs for all models presented in the article. Table VI presents the model coefficients and standard errors used to generate the predicted average attitudes. Table VII presents the model coefficients and standard errors in logits, which we converted to probabilities for the proportions of students above the 75% cutoff.

C. A

We are unaware of any single uniformly agreed to method for pooling the test results of the assumption checking for multilevel models when researchers use multiple imputation [108]. We performed the assumption checks using each imputed dataset. We present the results for the assumption checking using the pooled dataset made by averaging all of the imputed datasets. The pooled dataset on its own should not be used for checking the assumptions. We are, however, using it because our conclusions across all of the imputed models aligns with the results from the pooled data and to greatly simplify presentation of the assumption checking. To test the assumption of linearity, we plotted the residual variance against the fitted values, shown in Figs. 5 and 6. In our visual inspection of the figures we saw no obvious trends and concluded that the model met the assumption of linearity. To test for homogeneity of variance we created a box plot of the residuals across courses, shown in Figs. 5 and 6, and performed an ANOVA of the residuals across courses. A visual inspection of the box plot showed the courses' residuals had consistent medians and interquartile ranges and therefore met the assumption of homogeneity of variance. The ANOVA supported our visual check because it did not find a statistically significant difference ($p > 0.05$) in the variances across courses. Finally, we visually checked the assumption of normality of residuals using a qq plot of the observed and expected values, shown in Figs. 5 and 6. The small negative curvature in the qq plots indicated a small leftward skew in the residuals indicating there are more large negative residuals than a normal distribution would produce. This likely occurred because the data, as shown in the violin plots Fig. 4 tends to have a slight left (down) skew. Hence the model is overdriven by lower test scores. Gelman and Hill [125] point out that meeting the assumption of normally distributed residuals is of little importance to the regression line. The small skew in the residuals could have a very small effect on the standard errors. We expect that this skew had

TABLE VII. Hierarchical generalized linear model outputs for the pretest 75% cutoff models.

		Algebra-based		Calculus-based	
Race	Gender	Est.	SE	Est.	SE
Asian	Women	−2.13	0.20	−1.86	0.25
	Men	−1.50	0.21	−1.31	0.17
Black	Women	−2.61	0.43	−2.27	0.55
	Men	−1.41	0.40	−1.27	0.33
Hispanic	Women	−2.50	0.41	−1.80	0.37
	Men	−2.18	0.41	−1.93	0.28
White	Women	−1.58	0.11	−0.92	0.15
	Men	−1.38	0.11	−0.56	0.11
White Hispanic	Women	−2.26	0.30	−1.43	0.32
	Men	−1.67	0.29	−0.98	0.20

no effect on our conclusions for two reasons. First, the

[1] We use BIPOC instead of students of color to emphasize the historic and ongoing marginalization of Black and Indigenous peoples in the United States.

[2] American Physical Society, Diversity Statement, https://www.aps.org/policy/statements/08_2.cfm (2018).

[3] V. Sawtelle, E. Brewe, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, J. Res. Sci. Teach. **49**, 1096 (2012).

[4] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 020105 (2016).

[5] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a smog of bias, Phys. Rev. ST Phys. Educ. Res. **6**, 020112 (2010).

[6] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[7] B. Van Dusen and J. Nissen, Associations between learning assistants, passing introductory physics, and equity: A quantitative critical race theory investigation, Phys. Rev. Phys. Educ. Res. **16**, 010117 (2020).

[8] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?,

[32] 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).
[33] A. Madsen, S. B. McKagan, and E. C. Sayre, How

in the United States: A critical quantitative intersectionality analysis, Am. Educ. Res. J. **55**, 1268 (2018).
[65] C. D'

[101] S. Kanim and X. C. Cid, Demographics of physics education research, Phys. Rev. Phys. Educ. Res. **16**, 020106 (2020).
[102] M. Speidel, J. Drechsler, and S. Jolani, hmi: Hierarchical Multiple Imputation (2018), R package version 0.9.13.
[103] S. van Buuren and K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in R, J. Stat. Softw. **45**, 1 (2011).
[104] P. D. Allison, Missing Data (Sage Publications, Thousand Oaks, CA, 2001), Vol. 136.
[105] E. R. Buhi, P. Goodson, and T. B. Neilands, Out of sight, not out of mind: Strategies for handling missing data, Am. J. Health Behav. **32**, 83 (2008).
[106] J. L. Schafer, Multiple imputation: A primer, Stat. Methods Med. Res. **8**, 3 (1999).
[107] C. A. Manly and R. S. Wells, Reporting the use of multiple imputation for missing data in higher education research, Res. High. Educ. **56**, 397 (2015).
[108] 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).
[109] S. W. Raudenbush and A. S. Bryk, Hierarchical Linear Models: Applications and Data Analysis Methods (Sage, Thousand Oaks, CA, 2002), Vol. 1.
[110] S. Grund, A. Robitzsch, and O. Luedtke, mitml: Tools for Multiple Imputation in Multilevel Modeling (2018), R package version 0.3–6.
[111] D. Bates, M. Mächler, B. Bolker, and S. Walker, Fitting linear mixed-effects models using lme4, J. Stat. Softw. **67**, 1 (2015).