View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DigitalCommons@Florida International University

Florida International University
FIU Digital Commons

Department of Teaching and Learning

College of Arts, Sciences & Education

11-9-2015

Equity investigation of attitudinal shifts in introductory physics

Adrienne Traxler
Wright State University

Eric Brewe
Department of Teaching and Learning, Florida International University, ebrewe@fiu.edu

Follow this and additional works at: http://digitalcommons.fiu.edu/tl_fac

Part of the Education Commons

Recommended Citation
Traxler, Adrienne and Brewe, Eric, "Equity investigation of attitudinal shifts in introductory physics" (2015). *Department of Teaching and Learning.* Paper 4.
http://digitalcommons.fiu.edu/tl_fac/4

This work is brought to you for free and open access by the College of Arts, Sciences & Education at FIU Digital Commons. It has been accepted for inclusion in Department of Teaching and Learning by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

Equity investigation of attitudinal shifts in introductory physics

Adrienne Traxler*
Wright State University, 3640 Colonel Glenn Highway, Dayton, Ohio 45435, USA

Eric Brewe
Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA
(Received 20 September 2014; published 9 November 2015)

We report on seven years of attitudinal data using the Colorado Learning Attitudes about Science Survey from University Modeling Instruction (UMI) sections of introductory physics at Florida International University. University Modeling Instruction is a curricular and pedagogical transformation of introductory university physics that engages students in building and testing conceptual models in an integrated lab and lecture learning environment. This work expands upon previous studies that reported consistently positive attitude shifts in UMI courses; here, we disaggregate the data by gender and ethnicity to look for any disparities in the pattern of favorable shifts. We find that women and students from statistically underrepresented ethnic groups have gains that are comparable to those of men and students from well-represented ethnic groups on this attitudinal measure, and that this result holds even when interaction effects of gender and ethnicity are included. We conclude with suggestions for future work in UMI courses and for attitudinal equity investigations generally. We encourage researchers to expand their scope beyond simple performance gaps when considering equity concerns, and to avoid relying on a single measure to evaluate student success. Finally, we conjecture that students' social and academic networks are one means by which attitudinal and efficacy beliefs about the course are propagated.

DOI: 10.1103/PhysRevSTPER.11.020132 PACS numbers: 01.40.Fk, 01.40.gb

I. INTRODUCTION

The University Modeling Instruction curriculum (UMI; [1]) developed and studied at Florida International University (FIU) has produced an uncommon pattern of consistently positive shifts in student attitudes toward physics [2]. The case for studying student attitudes and epistemologies has been made at greater length elsewhere [2–4]; here, we will summarize those arguments, but largely take as a given that improving students' attitudes toward physics is one relevant dimension of success for a curriculum. However, education researchers must be cautious of overgeneralizing results, and one such overreach is to claim that a benefit is received by all students when in fact it only accrues to those from majority groups. FIU, a Hispanic-serving institution with a large fraction of women in the calculus-based modeling sections, provides an important opportunity to investigate this aspect of the UMI curriculum with a diverse student body. Section II discusses the context of gap-based analyses in education research, outlining some potential pitfalls of this approach and why we have chosen it here, and also summarizes some of the most relevant results on attitude surveys. Section III outlines the context of data collection and the research questions considered. Section IV summarizes our results, and Section V concludes with suggestions for future equity investigations of attitudinal or conceptual measures, cautioning to avoid forms of "gap gazing" that can further marginalize underrepresented groups.

II. BACKGROUND

A. Gaps analyses

Examination of performance differences, or looking for "gaps" between groups, is not without controversy in education research. As outlined by Gutiérrez [5] in mathematics and Danielsson [6] in physics, gaps analyses run the risk of essentializing student identities by overgeneralizing (e.g., "all women"). Gutiérrez argues that gap analyses often implicitly reinforce a deficit model in which students' differences are presumed to be the result of inadequacies in preparation, skill, or ability. Further, she argues, this frames students from different backgrounds in opposition with one another. Lubienski [7], on the other hand, contends that investigations of gaps are critically important to inform education policy and that it would be "irresponsible" to stop making gaps analyses. Following on Lubienski, we feel that it is not just valuable but essential for teachers and curriculum developers to question whether the benefits of instruction are distributed equitably among statistically underrepresented and majority student groups.

*adrienne.traxler@wright.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 3.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Some gap-based analyses, when thoughtfully conducted, have deepened our understanding of the mechanisms behind systemic performance differences on traditional academic measures. One key example is stereotype threat, originally uncovered when testing different framings of a difficult verbal test given to white and African American college students [8]. This landmark study and many following (for one review, see Ref. [9]) reveal a previously invisible barrier for women and students from statistically underrepresented racial and ethnic groups. Aware of negative stereotypes about their groups and invested in disproving them, these students face extra cognitive load from their awareness, and often show performance drops in the very subjects where they care the most [10].

Stereotype threat research has led to a richer understanding of how to frame classroom tasks in a manner that better supports all students. This work, including some in physics education research [11], would not have been possible without a willingness to investigate the causes of systematically observed performance differences between groups. Indeed, while Gutiérrez outlines pitfalls of gaps analyses, she also gives suggestions for avoiding them [5]. These suggestions include a greater focus on intervention work and on teaching and learning environments that support students from diverse racial, ethnic, and socioeconomic backgrounds. In the spirit of the second category, we focus our attention on data collected from University Modeling Instruction classes at FIU.

Rodriguez *et al.* [12] discuss three predominant models of equity in the context of physics education research: Equity of Fairness, Equity of Parity, and Equity of Individuality. Under the Equity of Fairness model, students from all populations should experience similar gains or losses. Equity of Fairness models would preserve preexisting gaps, such as the widely documented gender difference in Force Concept Inventory (FCI) pretest scores [13]. In the Equity of Parity model, students from one population might enter with lower scores on some measure, but all should leave with the same score distribution. Interventions striving for gap closing work from an Equity of Parity model. The study by Lorenzo, Crouch, and Mazur [14], showing a reduction or elimination in the FCI gender gap over a semester, is one example. Finally, Equity of Individuality investigations explicitly avoid group comparisons and instead focus on understanding individual excellence. An example of work in this category is a study by Goertzen, Brewe and Kramer [15] that uses case studies to examine several students' increasing levels of participation in the physics learning community at FIU, a large Hispanic-serving institution. Gap-based analyses are unable to speak to Equity of Individuality models, but may still provide important insight to Equity of Fairness or Equity of Parity questions. Research such as this paper, which explores differences in attitudinal shifts between groups, is relevant to Equity of Parity and Equity of Fairness models.

Previous work has outlined the epistemological goals of the UMI curriculum, which frames modeling as the key activity of scientists [1,16]. UMI classes have shown favorable student outcomes in conceptual understanding [17], in self-efficacy [18,19], in student social network measures [20], and in student attitudes towards physics [2] and engaging in physics [15]. We expand on the latter work here by examining whether these attitudinal gains are shared equally by women and by students from black, Hispanic, Native American, and Pacific Islander ethnicities. As of this paper's writing, all four ethnic groups are statistically underrepresented in the sciences and in physics, relative to the demographics of the United States population. Collected American Physical Society statistics highlight the situation, showing approximately 20% of physics degrees go to women, and less than 10% to African American, Hispanic, or Native American students [21]. In the text, we will adopt this language of "statistically underrepresented," to avoid possible deprecating connotations of "underrepresented minorities." The term also describes more accurately the prevalence of students of color in the broader United States, while reflecting that FIU is an uncommon example of a university where statistically underrepresented students are a majority.

B. Student attitudes

A variety of studies now document student attitudes in introductory university physics [3,22,23] and the effects of students' attitudes and epistemologies on their conceptual gains [22,24], use of content knowledge [25], and choice of courses and majors [24,26]. However, these results are not always reported through a lens of demographic factors. Some published CLASS results have shown more favorable pretest attitudes and shifts for men compared to women [23,27], while information about the effects of race or ethnicity is generally not available. This situation mirrors the literature on concept inventories, where gender gap results are often published [13], but ethnic representation is only rarely examined [17].

Despite a dearth of research on differential attitudes toward physics or science generally, the research on stereotype threat introduced above cautions us that attitudinal differences are very salient for students from statistically underrepresented groups. A serious long-term consequence of stereotype threat is the filtering effect it applies to participation: students from negatively stereotyped groups, over time, often disidentify with the threatened area [28]. A negative shift in attitudes toward a subject can be an important self-preservation tactic for students who are threatened by stereotype. By devaluing the domain, they minimize the risks that the stereotype-predicted poor performance would otherwise pose to their self-image. While the CLASS does not measure physics identity directly, gender patterns in published results are troubling [23,27]. Additionally, a study asking students to

report both their own and a scientist's expected answers on the CLASS found that women were equally or better able to select the "scientist" response, but the gap with their own answers was greater [29]. Because prolonged stereotype threat affects domain identification, differentially lower initial attitudes or negative attitudinal shifts may be an important warning to instructors of disengagement in students from threatened groups.

Research from the University of Colorado has shown that initial (pre-university instruction) student attitudes are strongly correlated with pursuing a physics major [26]. It is possible that strong positive shifts may also show some similar effect on recruiting students to the major. This question remains open, in part, because demonstrating consistently positive shifts has been a substantial task. However, at a more fine-grained scale, a positive shift in attitudes toward physics learning has been linked with more central membership in the physics community in FIU's rapidly growing physics major population [15].

We have ample motivation to examine patterns of positive attitudinal shifts, as potential signals of growing student investment and participation in physics. However, to accurately report promising findings, we must also ask whether any such benefits are equally received by all groups of interest. In this paper, we investigate precourse to postcourse attitude scores and shifts for students in calculus-based Physics I (mechanics) courses. From previous work, we know that the University Modeling Instruction courses are equitable by the Equity of Fairness model for Force Concept Inventory gains by ethnicity, but not by gender [17]. In other words, student gains were independent of their ethnic representation, but a gender gap widened over the semester (so not all student groups experienced equal gains). Here we extend the equity question to attitudinal shifts. This investigation contributes to the knowledge base on impacts of student attitudes by first exploring differential attitudes across statistically underrepresented student groups, and then by asking how instruction shifts student attitudes among these groups.

III. METHODS

FIU is a large, minority-serving institution (54 000 students, 61% Hispanic, 13% black, in Spring 2014) with a primarily commuter student body. Over the past ten years, the Physics Education Research Group has guided a series of structural reforms in the introductory physics courses, including the addition of University Modeling Instruction sections of the calculus-based sequence. The data presented in this paper are drawn from introductory physics I courses and were collected from the Fall 2007 to Fall 2013 semesters. Table I shows the demographics of the student sample. The gender ratio is much closer to parity in UMI sections than in traditional lecture physics courses at FIU, while the distribution of students' ethnic representation is very similar between the two course formats. Because of the popularity of the Modeling Instruction sections, students are admitted by lottery.

TABLE I. Demographics of University Modeling Instruction sections in the sample ($N = 264$). Counts and percentages are given for gender and for ethnic representation, grouping by statistically well- or overrepresented (SR) and statistically underrepresented (SUR).

		$N[\%]$	
		Female	Male
Ethnicity (SR)	Asian	4 [1.5]	12 [4.5]
	White	6 [2.3]	15 [5.7]
	Black	11 [4.2]	11 [4.2]
Ethnicity (SUR)	Hispanic	115 [43.6]	88 [33.3]
	Native Amer.	0 [0]	1 [0.4]
	Pacif. Islander	0 [0]	1 [0.4]
Total		136 [51.5]	128 [48.5]

The Colorado Learning Attitudes about Science Survey [23] is a 42-item Likert-scale instrument, where students select their level of agreement or disagreement with statements about physics. Answers are compared to an expert response key to give a "percentage favorable" score. The CLASS was administered on paper at the beginning and end of each term and filtered for matched student responses, which are necessary to calculate shifts. We look for pretest, post-test, and shift differences between students who are statistically well- or overrepresented in physics (male, Asian, and white students) and those who belong to statistically underrepresented groups (female, black, Hispanic, Native American, and Pacific Islander students). Several ethnicities are represented by only a few students in our data set. To alleviate the statistical difficulties of small sample sizes, and to look for broad differences by representation rather than fine-grained comparisons between groups, the ethnicity component of the analysis will only distinguish between statistically well- or overrepresented (SR) and statistically underrepresented (SUR) categories.

We seek to answer two research questions:

(1) To what extent does gender or ethnic representation influence students' percentage of expertlike CLASS responses in University Modeling Instruction?

(2) To what extent is there an interaction between gender and ethnic representation?

To address the first question, we disaggregate student pretests, post-tests, and shifts in percentage favorable responses on the CLASS. In addition to checking for statistically significant differences in these values between groups, we follow Rodriguez *et al.* [12] in looking for significant effect sizes. We measure effect size using Cohen's d [30]:

$$d = \frac{\mu_2 - \mu_1}{\sigma_{\text{pooled}}}. \tag{1}$$

Here, μ_1 and μ_2 are the means of two groups to be compared (e.g., mean precourse scores for men and women, $\mu_{\mathrm{M,pre}}$ and $\mu_{\mathrm{F,pre}}$), and σ_{pooled} is the pooled standard deviation of the two groups. The effect size provides an indicator of "practical significance," and thus serves as a necessary accompaniment to statistical significance when reporting claims about gaps between groups [12].

The second question occurs because the intersection of gender and racial or ethnic identity is known to pose additional challenges for women of color in the sciences [31]. To address this point, we use a linear regression model including an interaction term for gender and ethnicity, and investigate whether it explains a significant amount of the variance in postcourse attitudes.

IV. RESULTS

Figure 1 shows the significant and positive differences between pre- and post-course responses in the modeling classes. Disaggregating by gender and by ethnic representation, we see that all subgroups show significant positive shifts. On close examination of the histograms for pre- and post-course percentage favorable scores (not pictured), some ceiling effect may exist in the post scores. However, it is not notably more pronounced for any of the subgroups. Figure 2 elaborates on the disaggregated results by showing percentage favorable shifts for all students, by gender, and by ethnic representation. Average shifts do not vary by ethnic representation, but women do have a significantly higher average shift than men (see Table II in the Appendix for values and standard errors).

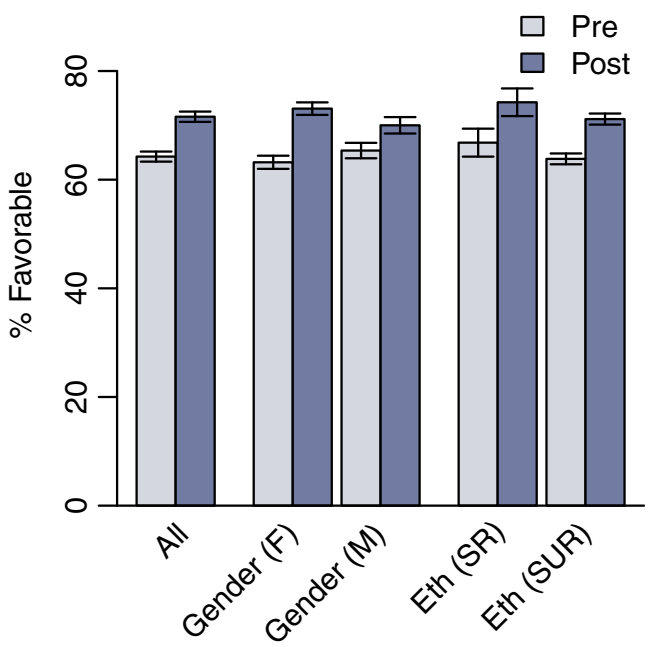

FIG. 1. Overall favorable average CLASS scores for Modeling Instruction sections. Bars show standard error of the mean.

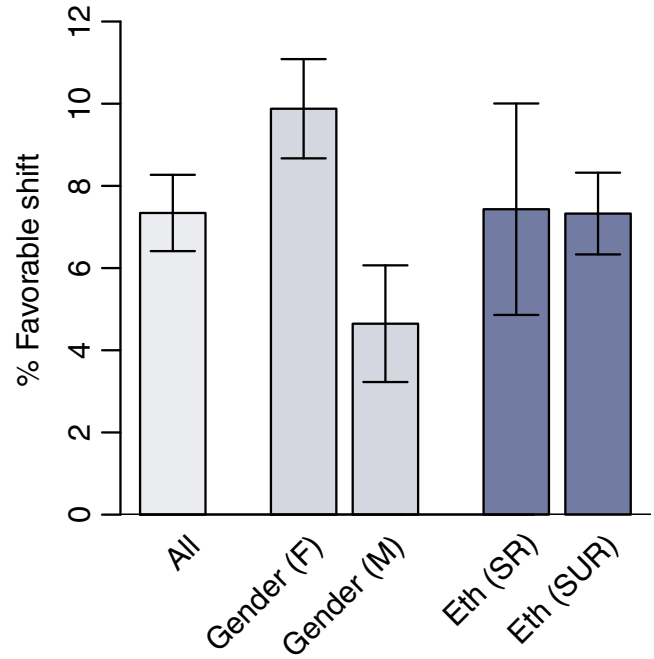

FIG. 2. Average shift in overall favorable percentage CLASS score, with bars showing standard error of the mean. Shifts are shown overall and then disaggregated by gender (male or female) and ethnic representation (statistically represented or statistically underrepresented).

Figure 3 shows the effect sizes, Cohen's d, of group differences on pretest and post-test. We see that for both gender and ethnicity, on precourse and postcourse administrations of the CLASS, the effect sizes of the differences are small ($|d| \lesssim 0.2$) and the error bars span zero. This overlap indicates that there is no meaningful difference between the pre- and postcourse means for men compared to women, or statistically represented compared to statistically underrepresented ethnicities. As advocated by Rodriguez *et al.* [12], we find that examination of effect sizes adds nuance beyond that provided by null hypothesis significance testing. Figure 1 shows that women's precourse averages were somewhat lower and their postcourse averages somewhat higher compared to men, producing a statistically significant higher average shift (Fig. 2). The nonsignificant effect sizes in Fig. 3 for pre- or postcourse gender differences caution that this shift does not indicate a

TABLE II. Average precourse, postcourse, and shifts in percentage favorable responses ± standard error of the mean.

Group	N	Pre	Post	Shift
All students	264	64.2 ± 0.9	71.6 ± 0.9	7.3 ± 0.9
Female	136	63.2 ± 1.2	73.1 ± 1.2	9.9 ± 1.2
Male	128	65.4 ± 1.4	70.0 ± 1.5	4.6 ± 1.4
SR	37	66.8 ± 2.6	74.2 ± 2.5	7.4 ± 2.6
SUR	227	63.8 ± 1.0	71.2 ± 1.0	7.3 ± 1.0

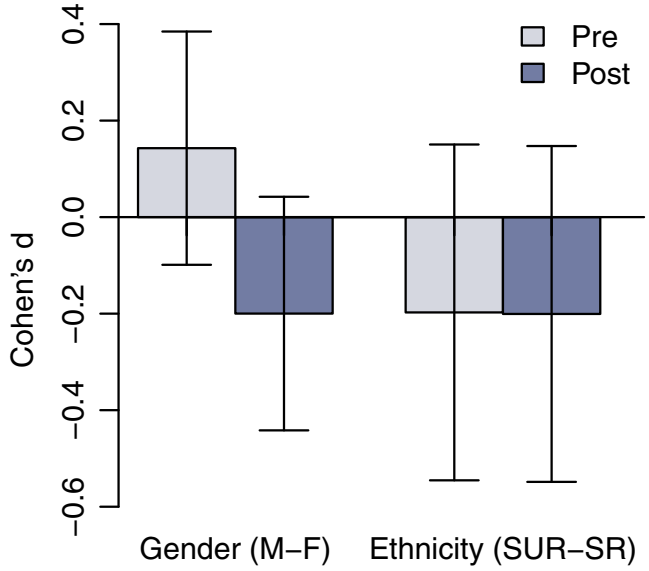

FIG. 3. Effect sizes of disaggregated groups. In all cases, differences between group means are at or below the threshold for small effect size, with error bars crossing the axis indicating no meaningful effect.

practically significant difference in the distributions by gender. One final caveat comes from the sample sizes, which are comparable for gender but uneven for ethnic representation (with relatively few statistically represented students at FIU). As per Cohen's notes on statistical power [30], there is some risk that a small effect for gender or a medium-size effect for ethnic representation may have been missed due to sample size constraints. Accordingly, the effect size findings contextualize Figs. 1 and 2, but will be more robust as future data is accumulated. Section V revisits effect size in the context of equity models.

Finally, to check for possible interactions of gender and ethnicity that might be overlooked when considering each factor individually, we use a linear regression model:

$$\text{Post} \sim \text{Pre} + \text{Gender} + \text{EthRep} + \text{Gender} \times \text{EthRep}. \quad (2)$$

Here, Post and Pre represent the overall percent favorable scores, Gender is coded as F or M [32], and EthRep is coded SR or SUR for statistically represented or underrepresented ethnic groups, respectively.

Fitting this model to the sample of 264 students, we find that only the coefficient for Pre is significant: $\beta_{\text{Pre}} = 0.57$, $95\%\text{CI} = (0.46, 0.67)$, $p < 0.01$. For the full model, $R^2 = 0.32$, indicating that substantial variance remains unexplained. Neither gender nor ethnic representation, or the interaction between them, were significant predictors of post-course expertlike beliefs once a student's precourse beliefs were accounted for. This result supports the non-significant effect sizes found above, and clarifies that there is no detectable gender-ethnicity interaction that was hidden by splitting the data along those categories.

V. DISCUSSION AND CONCLUSIONS

Previous studies of student conceptual gains in introductory physics have pointed to a disparity in scores between male and female students [13,17]. Results vary on whether these gaps persist in reform-based classes, where various features of the learning environment might be expected to support traditionally marginalized students. Although there is important debate about the degree to which gap gazing is useful or appropriate in education research, a gender or ethnicity-divided difference in gains is troubling because it suggests that not all students are receiving the claimed benefits of reform efforts.

In the attitudinal study reported here, the picture is somewhat different than for conceptual measures. Returning to our research questions:

1. To what extent does gender or ethnic representation influence students' percentage of expertlike CLASS responses in University Modeling Instruction? There is no evidence that female students, or those from statistically underrepresented ethnicities, have either lower or higher precourse, postcourse, or shifts in percentage of favorable beliefs on the CLASS. Closer examination of the score distributions does show some evidence of a ceiling effect on the postcourse CLASS, but it does not appear that the effect is stronger for men or students from SR ethnicities (which, if it had been the case, might artificially suppress a gap). It would be very useful to disaggregate the scores by gender and ethnicity for a broader sample of classes, where high pretest scores are less prevalent, and for non-modeling courses (more on this below).

2. To what extent is there an interaction between gender and ethnic representation? In a linear model of postcourse attitude scores where gender, ethnicity, and their interaction are included, none of these coefficients are statistically significant. Only students' precourse attitudes are a significant predictor in the model, and even with this inclusion the model only accounts for 32% of the total variance in postcourse attitude scores. So far as we can detect with these data women from statistically underrepresented ethnicities have a similar pre- and postcourse attitude profile as their peers in other groups.

Revisiting the two models of equity discussed in Section II, the modeling classes are supportive of student attitudes in the Equity of Fairness sense, where all groups show similar gains. No precourse differences in distribution were detected, nor did traditional majority groups show disproportionate gains, so Modeling Instruction is also supportive of student attitudes by the Equity of Parity model.

As noted above, one possible explanation for FCI gender gaps is stereotype threat, which is known to depress the performance of women and students from underrepresented ethnic groups on many academic tasks. However, a key component of the threat is perceived risk of doing badly on a task where one will be judged. Three features of the current study—some of which may be idiosyncratic to the

FIU context—are worth discussing as pertaining to stereotype threat. First, there may be some mediating effect on stereotype threat when students who typically are impacted by stereotype threat, such as statistically underrepresented groups, are in the majority. This explanation fails to account for the lack of precourse gender differences in our sample, and in this study we do not make such a claim. A second point is that, even when statistically underrepresented students are in the majority—as is the case at FIU—the instructor still holds a position of power whereby they can make evaluative decisions. However, an attitudinal survey, where students are asked to rate their beliefs rather than to choose one correct answer, may be perceived as a less failure-prone task and thus not trigger the threat. Finally, the precourse attitudes for students entering the modeling classes are already very favorable, above lecture students at the same institution, and at the high end of typical precourse scores reported for the CLASS (Adams *et al.* [23], Tables V and VIII).

Possible explanations include greater student buy-in at the beginning of the semester, as students must apply and be selected by lottery due to the popularity of the UMI sections. A related hypothesis is that the same informal network of peers that passes information about the course may also confer a higher expectation of success, leading to a self-efficacy boost that registers on the CLASS. To help account for the first possibility, a more comprehensive attitude survey of lecture students in the same cohort would be useful: tracking students who unsuccessfully applied to modeling sections, and comparing their CLASS precourse scores with those who found seats in modeling, could detect whether the UMI classes somehow attract more "physics people." Observationally, however, this is somewhat unlikely, as many UMI students are on premedical paths and have no initial interest in a physics career.

Returning to the question of gap gazing, looking for performance differences between groups should be done carefully, because it risks problematizing already marginalized students. But until the field of physics accurately reflects the diverse talents of the population, and until effects such as stereotype threat are no longer detectable, it is important for education researchers to address whether their reforms truly are for all students. Building on this awareness, a constructive way to address the problems of underrepresentation in science is to examine successful curricula and learning environments so that lessons may be drawn from positive examples.

In this work, we have examined the favorable attitudinal shifts reported in UMI courses, asking whether they are equitable among students of different genders and ethnicities. We find that they are, and somewhat surprisingly, that this is true even on a precourse attitude survey where more negative attitudes have been reported for women in other studies. While it would be unreasonable to attribute this precourse parity to the UMI curriculum, it suggests that a fruitful dimension for research to expand is beyond the boundaries of classroom pre- and post-tests, to investigate the learning networks and communities that may transmit information and expectations to future students. The results reported here, taken together with previous FCI and odds of success comparisons for the same courses [17], also caution against taking any one test score—attitudinal, conceptual, or otherwise—as the solitary measure of a student group. Multiple measures of success are needed to understand, measure, and value the many things that students learn in physics courses.

ACKNOWLEDGMENTS

This manuscript is based upon work supported by the Howard Hughes Medical Institute's Science Education Grant No. 52006924 and by the National Science Foundation under Grant No. 0802184 and DUE 1140706. The authors would like to thank the students who participated in this study and the FIU Physics Education Research Group for their insight.

APPENDIX: FIGURE DATA

Table II gives the values used to generate Figs. 1 and 2 for completeness and to facilitate comparison with other studies.

[1] E. Brewe, Modeling theory applied: Modeling Instruction in introductory physics, Am. J. Phys. **76**, 1155 (2008).

[2] E. Brewe, A. Traxler, J. de la Garza, and L. H. Kramer, Extending positive CLASS results across multiple instructors and multiple classes of Modeling Instruction, Phys. Rev. ST Phys. Educ. Res. **9**, 020116 (2013).

[3] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics, Am. J. Phys. **66**, 212 (1998).

[4] National Research Council, *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, edited by S. R. Singer, N. R. Nielsen, and H. A. Schweingruber (National Academies Press, Washington, D.C., 2012).

[5] R. Gutiérrez, A "gap-gazing" fetish in mathematics education? Problematizing research on the achievement gap, *J. Res. Math. Educ.* ***39****, 357 (2008).*

[6] A. T. Danielsson, Gender in physics education research: A review and a look forward, in *Never Mind the Gap! Gendering Science in Transgressive Encounters*, Crossroads of Knowledge, Vol. 15, edited by M. Blomqvist and E. Ehnsmyr (University Printers, Uppsala, Uppsala, Sweden, 2010), pp. 65–83.

[7] S. T. Lubienski, On "gap-gazing" in mathematics education: The need for gaps analyses, *J. Res. Math. Educ.* ***39****, 350 (2008)*.

[8] C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African Americans, J. Pers. Soc. Psychol. **69**, 797 (1995).

[9] J. Aronson and C. M. Steele, Stereotypes and the fragility of human competence, motivation, and self-concept, in *Handbook of Competence and Motivation*, edited by A. J. Elliot and C. S. Dweck (Guilford Press, New York, 2005), p. 436.

[10] J. Aronson, M. J. Lustina, C. Good, K. Keough, C. M. Steele, and J. Brown, When white men can't do math: Necessary and sufficient factors in stereotype threat, J. Exp. Soc. Psychol. **35**, 29 (1999).

[11] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation, Science **330**, 1234 (2010).

[12] I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer, Impact of equity models and statistical measures on interpretations of educational reform, Phys. Rev. ST Phys. Educ. Res. **8**, 020103 (2012).

[13] 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? Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).

[14] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[15] R. M. Goertzen, E. Brewe, and L. Kramer, Expanded markers of success in introductory university physics, Int. J. Sci. Educ. **35**, 262 (2013).

[16] D. Hestenes, Toward a modeling theory of physics instruction, Am. J. Phys. **55**, 440 (1987).

[17] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in Modeling Instruction in introductory university physics, Phys. Rev. ST Phys. Educ. Res. **6**, 010106 (2010).

[18] V. Sawtelle, E. Brewe, and L. H. Kramer, Positive impacts of Modeling Instruction on self-efficacy, in *2010 Physics Education Research Conference*, AIP Conf. Proc. No. 1289, edited by C. Singh, M. Sabella, and S. Rebello (AIP, New York, 2010), pp. 289–292.

[19] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, Creating opportunities to influence self-efficacy through Modeling Instruction, in *2011 Physics Education Research Conference*, AIP Conf. Proc. No. 1413, edited by N. S. Rebello, P. V. Engelhardt, and C. Singh (AIP, New York, 2012), pp. 339–342.

[20] E. Brewe, L. H. Kramer, and G. E. O'Brien, Changing participation through formation of student learning communities, in *2010 Physics Education Research Conference*, AIP Conf. Proc. No. 1289, edited by C. Singh, M. Sabella, and Sanjay Rebello (AIP, New York, 2010), pp. 85–88.

[21] http://www.aps.org/programs/education/statistics/, sourced from the IPEDS Completion Survey.

[22] I. Halloun, Views about science and physics achievement: The VASS story, in *The changing role of physics departments in modern universities*, AIP Conf. Proc. No. 399 (AIP, College Park, MD, 1997), pp. 605–614.

[23] 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).

[24] 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, in *2004 Physics Education Research Conference*, AIP Conf. Proc. No. 790, edited by J. Marx, P. Heron, and S. V. Franklin (AIP, Sacramento, CA, 2005), pp. 61–64.

[25] L. Lising and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. **73**, 372 (2005).

[26] K. K. Perkins and M. Gratny, Who becomes a physics major? A long-term longitudinal study examining the roles of pre-college beliefs about physics and learning physics, interest, and academic achievement, in *2010 Physics Education Research Conference*, AIP Conf. Proc. No. 1289, edited by C. Singh, M. Sabella, and S. Rebello (AIP, Portland, Oregon, 2010), pp. 253–256.

[27] 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).

[28] C. M. Steele, A threat in the air: How stereotypes shape intellectual identity and performance, Am. Psychol. **52**, 613 (1997).

[29] K. E. Gray, W. K. Adams, C. E. Wieman, and K. K. Perkins, Students know what physicists believe, but they don't agree: A study using the CLASS survey, Phys. Rev. ST Phys. Educ. Res. **4**, 020106 (2008).

[30] J. Cohen, A power primer, Psychol. Bull. **112**, 155 (1992).

[31] M. Ong, C. Wright, L. L. Espinosa, and G. Orfield, Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering,and mathematics, Harv. Educ. Rev. **81**, 172 (2011).

[32] We acknowledge that this strict gender binary is a simplification, but our demographic data come from the university database, which presently only allows these two choices.