Racial and ethnic bias in the Force Concept Inventory

Rachel Henderson[1] and John Stewart[1]

[1]*Department of Physics and Astronomy, West Virginia University, 135 Willey St, Morgantown, WV, 26506*

Gender gaps on the various physics concept inventories have been extensively studied; however, little work has been done to explore racial or ethnic biases. In this work, we will present student averages on the Force Concept Inventory (FCI) for African-American (n=85), Caucasian (n=1665), and Hispanic (n=82) students. Hierarchical linear regression was used to explore the effects of gender and race/ethnicity on FCI posttest scores. As expected from the literature, a gender gap of 8% was found in the FCI posttest scores with male students outperforming female students. Differences in FCI posttest performance were also measured with Caucasian students outperforming African-American students (14%) and Hispanic students (6%). After controlling for course performance measured by physics grade, the differences narrowed somewhat; Caucasian students outperformed both African-American students (9%) and Hispanic students (4%) on the FCI posttest. No significant interaction between gender and race/ethnicity was found; the gender gap identified in the majority population was shared equally by all students.

I. INTRODUCTION

Commonly used conceptual mechanics evaluations, such as the Force Concept Inventory (FCI) [1] and the Force and Motion Conceptual Evaluation (FMCE) [2], have consistently demonstrated a difference in performance between men and women. Research into the "gender gap" was summarized by Madsen, McKagan, and Sayre [3] who reported that, on average, male students outperformed female students by 12% in posttest scores on the mechanics conceptual inventories and by 8.5% on electricity and magnetism conceptual inventories. Many factors that may influence the gender gap have been explored such as background and preparation [4, 5], teaching method or instructor [6–8], and sociocultural factors [9, 10]. Kost, Pollock and Finklestein demonstrated that differences in FMCE pretest score accounted for a substantial fraction of the gender gap in the FCME posttest [4]. Interactive engagement has been demonstrated as an important factor in learning conceptual physics [11, 12]. While some studies have shown that interactive engagement methods are productive in reducing the gender gap [7], this result has not be reproduced in all studies [6]. For a more complete exploration of gender and physics, see Traxler *et al.* [13].

While the gender gap in conceptual physics has been extensively studied, little research has explored whether these gaps extend to other underrepresented populations such as students of race or ethnicity different from the majority Caucasian population. In a recent commentary, Scherr lamented the unbalanced focus of PER on gender when many other populations are substantially underrepresented in physics [14]. A few studies have examined differences in posttest scores by race and ethnicity. Kost, Pollock and Finklestein found that ethnicity was not a significant factor in predicting FMCE posttest scores; however, they warned that this could be due to the small number of students of non-majority ethnicity in their sample [4]. In another study, Kost-Smith, Pollock, and Finklestein added ethnicity to their model when predicting Brief Electricity and Magnetism Assessment (BEMA) [15] posttest scores. They also found that ethnicity did not ex-

plain additional variability in posttest scores [5]. Hazari, Tai, and Sadler showed that there was a difference in introductory course performance, measured by grade, between Caucasian, African-American, and Hispanic students [16]. To the authors' knowledge, the validity and reliability of the FCI has not been explored for any racial or ethnic groups other than the majority Caucasian students.

The underrepresentation of women and minorities within Science, Technology, Engineering, and Mathematics (STEM) fields is a serious national concern. According the National Science Foundation, in 2012, 60% of bachelor's degrees awarded in science and engineering went to Caucasian students, 8.4% to African-American students, and 9.9% to Hispanic or Latino students [17]. Also in 2012, out of 5,557 awarded bachelor's degrees in physics, only 142 and 314 degrees in were given to African-American and Hispanic students respectively [17]. Although bachelor's degrees in STEM awarded to African-Americans grew from 1995 to 2004, an inverse relationship still exists between degree level and number awarded to African-Americans [18]. Chapa and De La Rosa reviewed the underrepresentation of Latino students in higher education. Although the Latino population has a smaller college participation rate than the Caucasian population, by 2040, the number of Latinos that enroll in college will increase from 1 million to almost 2 million [19].

African-American and Hispanic men and women are not as likely to attend and complete college as Caucasian students. According to the *ACT Policy Report*, of the African-American and Hispanic students enrolled in a four-year college, 41% completed a degree within six years at the same institution compared to 59% of Caucasian students [20]. Braxton, Milem, and Sullivan demonstrated that race had a significant direct effect on intent to re-enroll in the upcoming semester [21]. Ishitani analyzed departure rates of transfer students; minority transfer students were 68% more likely to depart than their Caucasian transfer counterparts [22]. Toven-Lindsey *et al.* demonstrated that minority students who enrolled in an academic support program were more likely to persist in a science major [23]. Sociocultural factors that have been shown to be important for retaining women in STEM

have also been investigated for underrepresented minorities [24, 25].

This study seeks to add to the sparse literature on the differences in conceptual physics performance between Caucasian, African-American, and Hispanic introductory physics students by answering the following research questions: *RQ1: Are there differences in FCI posttest scores between male and female students? RQ2: Are there differences in course grades or FCI posttest scores between students of different race and ethnicity? RQ3: If a gender difference exists in FCI posttest scores, is this gender difference the same for students of all races and ethnicities?*

II. METHODS

This research was conducted in the first-semester, calculus-based mechanics course at a midwestern land-grant university serving approximately 25,000 students in the United States. This course consisted of two 50-minute lectures and two two-hour laboratory sessions each week. The class was overseen by the same lead instructor who taught most lecture sections (75%) during the period studied. Students completed homework assignments, lecture and laboratory quizzes, and in-semester examinations. The FCI was given both pre- and post-instruction. For the pretest, credit was given for a good faith effort; however, the posttest was given as a part of the last in-semester examination.

From the fall 2006 semester to spring 2012 semester, 3,273 students completed the course for a grade. Of these students, 3,237 completed the FCI posttest. Students were also asked to self-report race and ethnicity as part of a survey administered at a different time than the FCI. Of the 3,237 students who completed the posttest, 2,038 reported their race or ethnicity. Race and ethnicity was collected as "Caucasian" ($n = 1665$), "African-American" ($n = 85$), "Asian/Pacific Islander" ($n = 124$), "Hispanic" ($n = 82$), and "Other" ($n = 82$). Only students that reported African-American (4%), Caucasian (82%), or Hispanic (4%) were included in this study, leaving a sample size of 1,832 students (74% male students). For all analyses, gender was coded with female as 0 and male as 1 and Caucasian students were coded as the reference group. Course letter grades were measured on a four-point scale with A=4 and F=0. To explore the differences in FCI posttest scores between the various racial and ethnic groups ANOVA and hierarchical linear regression were employed.

III. RESULTS

Table I summarizes FCI posttest scores and course letter grades; FCI scores are presented as a percentage. Significance levels were Bonferroni corrected to adjust for inflation of Type I error. Effect size was characterized by Cohen's *d*; Cohen's convention for effect size identifies $d = 0.2$ as a

small effect, $d = 0.5$ as a medium effect, and $d = 0.8$ as a large effect.

TABLE I. Course letter grades and FCI posttest averages. Letter grades were measured on a four-point scale and FCI posttest averages are reported as percentages.

	Male Students		Female Students		
	N	$M \pm SD$	N	$M \pm SD$	d
African-American ($n = 85$)					
Physics Grade	61	2.9 ± 1.1	24	3.0 ± 0.7	0.11
FCI Posttest	61	63 ± 19	24	58 ± 13	0.32
Caucasian ($n = 1665$)					
Physics Grade	1241	3.4 ± 0.8	424	3.5 ± 0.7	0.17
FCI Posttest	1241	78 ± 15	424	70 ± 15	0.52
Hispanic ($n = 82$)					
Physics Grade	54	3.3 ± 0.8	28	3.1 ± 0.9	0.23
FCI Posttest	54	72 ± 15	28	66 ± 12	0.40

Differences between the students of different race and ethnicity were analyzed using a one-way between subjects ANOVA. Results showed that there were significant differences in both letter grade [$F(2, 1829) = 16.67, p < 0.001$] and FCI posttest score [$F(2, 1829) = 36.86, p < 0.001$].

A posthoc analysis showed that there were significant differences in course grade between Caucasian and African-American students ($p < 0.01$) with a medium effect size of $d = 0.63$ and between Hispanic and African-American students ($p < 0.05$) with a small effect size of $d = 0.43$. There was no significant difference in course grade between Caucasian and Hispanic students. For the FCI posttest score, posthoc analysis showed significant differences between all races and ethnicities ($ps < 0.01$). There was a large effect size between Caucasian and African-American students ($d = 0.90$), a medium effect size between Hispanic and African-American students ($d = 0.52$), and a small effect size between Caucasian and Hispanic students ($d = 0.36$).

Within each racial and ethnic group, differences by gender were analyzed using *t*-tests. For African-American students and Hispanic students there were no significant gender differences in either course grade or FCI posttest score. However, for Caucasian students, there was a significant gender difference in course grade [$t(814) = 3.13, p < 0.05, d = 0.17$] and FCI posttest score [$t(723) = 9.10, p < 0.001, d = 0.52$]. While significant gender differences were only measured for Caucasian students, the size of the differences were very similar between the three groups. An increase in sample size for African-American and Hispanic students may lead to the findings of significant gender differences.

To more thoroughly explore how the FCI posttest percentage was related to gender, race, and ethnicity, hierarchical linear regression was employed. Table II presents the results of this analysis. Hierarchical linear regression is a technique to determine if the addition of an independent variable sig-

nificantly improves the proportion of explained variance in the dependent variable. This technique adds one additional independent variable in a step-wise fashion and assesses the differences between the successive models.

TABLE II. Hierarchical linear regression analysis predicting FCI posttest percentage. Female was coded as 0 and male as 1. B is the regression coefficient, SE the standard error, and β the regression coefficient normalizing the posttest percentage. Superscript "a" denotes $p < 0.05$, "b" denotes $p < 0.01$, and "c" denotes $p < 0.001$. The R^2_{adj} significance levels indicate the significance of the improved fit of the model over the model in which it is nested.

		Variables	B	SE	β	R^2_{adj}
Model 1		Physics Grade	10.18^c	0.39	0.52^c	$.27^c$
Model 2		Gender	7.76^c	0.83	0.49^c	$.05^c$
Model 3		African-Amer.	-14.03^c	1.73	-0.88^c	$.04^c$
		Hispanic	-5.64^b	1.76	-0.36^b	
Model 4		Physics Grade	9.84^c	0.39	0.50^c	$.28^c$
		African-Amer.	-9.07^c	1.51	-0.57^c	
		Hispanic	-4.48^b	1.52	-0.28^b	
Model 5	Step 1	Physics Grade	10.46^c	0.38	0.53^c	$.33^c$
		Gender	8.88^c	0.70	0.56^c	
	Step 2	Physics Grade	10.14^c	0.38	0.51^c	$.34^c$
		Gender	8.73^c	0.69	0.55^c	
		African-Amer.	-8.68^c	1.45	-0.55^c	
		Hispanic	-3.68^a	1.46	-0.23^a	
	Step 3	Physics Grade	10.16^c	0.38	0.52^c	.34
		Gender	9.13^c	0.73	0.58^c	
		African-Amer.	-6.88^c	2.71	-0.43^a	
		Hispanic	-0.09	2.52	-0.01	
		Afr.-Amer. $\times$ Gend.	-2.47	3.19	-0.16	
		Hisp. $\times$ Gend.	-5.40	3.09	-0.34	

In Model 1, course performance, measured by overall physics grade, was a significant predictor of FCI posttest average. This independent variable alone explained 27% of the variability in posttest scores.

Model 2 explored the relationship of gender and FCI posttest averages. There was a significant gender gap with male students outperforming female students by 7.76%; however, only 5% of the variance in FCI posttest average was explained by gender.

Model 3 examined the relationship of race and ethnicity with posttest average. This model compares African-American and Hispanic students to Caucasian students; Caucasian students form the baseline for the regression and the regression coefficient measures the change with respect to this baseline. On average, African-American students (14.03%) and Hispanic students (5.64%) scored lower on the FCI posttest than Caucasian students. Race and ethnicity explained only 4% of the variability in FCI posttest score.

Model 4 explored the effect of race and ethnicity controlling for course performance. After controlling for physics

grade, there were significant differences in FCI posttest scores between students of different race and ethnicity; African-American students (9.07%) and Hispanic students (4.48%) performed more weakly than Caucasian students. These differences were smaller than the uncorrected differences in Model 3. As such, some but not all of the differences in posttest scores for these students were explained by overall class performance.

Model 5 examined the relationship between gender, race, and ethnicity while controlling for course grade. Model 5–Step 1 identified a significant overall gender gap (8.88%) on the FCI posttest controlling for overall course performance measured by physics grade. Model 5–Step 1 significantly improved model fit ($p < 0.001$) over Model 1 and explained 33% of the variability in posttest scores. Model 5–Step 2 added race and ethnicity to the model. There was still a significant gender difference in posttest scores (8.73%) as well as significant differences between each racial and ethnic group. Controlling for course grade, African-American students still scored significantly lower than Caucasian students (8.68%) while Hispanic students also scored significantly lower than Caucasian students (3.68%). Although adding race and ethnicity to the model explained only an additional 1% of variability in posttest average, Model 5–Step 2 was a significantly better model than Model 5–Step 1 ($p < 0.001$).

Model 5–Step 3 introduced interactions between gender, race, and ethnicity. This model did not explain significantly more variability in posttest average and was not a significantly better model when compared to Model 5–Step 2. The interaction terms in this model were also not statistically significant. Although an overall gender gap exists in the FCI posttest score, this gender gap was the same for African-American, Caucasian, and Hispanic students correcting for course performance.

IV. DISCUSSION

RQ1: Are there differences in FCI posttest scores between male and female students? A gender gap of 7.76% was found in FCI posttest scores with men outperforming women (Table II Model 2). When controlling for course performance this difference in FCI posttest scores between male and female students was still found to be significant and relatively unchanged (8.88%) (Table II Model 5 – Step 1). Although this gender gap was lower than the overall average of 12% reported by Madsen, McKagen, and Sayre [3], the result was in the range of gender differences in FCI posttest scores in their review.

RQ2: Are there differences in course grades or FCI posttest scores between students of different race and ethnicity? The differences in course grades between Caucasian, African-American, and Hispanic students were similar to those presented previously [16]. Differences were also identified in FCI posttest scores. Controlling for course performance, these differences persisted but narrowed. Controlling

for course grade, a 9.07% difference was measured between African-American students and Caucasian students and a difference of 4.48% between Hispanic students and Caucasian students (Table II Model 4). This study detected not only a gender gap in the FCI but also racial and ethnic differences not previously reported [4, 5]. Only some of the racial and ethnic gaps were explained by course grades while none of the gender gap was explained by grades.

RQ3: If a gender difference exists in FCI posttest scores, is this gender difference the same for students of all races and ethnicities? Comparison of Model 4 and Model 5–Step 2 in Table II showed that there was both an overall effect of gender (8.73%) and of race/ethnicity [African-American 8.68%; Hispanic 3.68%] which coexisted. The failure to find interactions between gender and race/ethnicity in Model 5–Step 3 suggests the gender gap is consistent across students of all races and ethnicities. The gender gap was neither localized in the majority Caucasian population, nor more or less severe for African-American or Hispanic students.

V. IMPLICATIONS AND LIMITATIONS

The effect of gender was fairly orthogonal to the effect of race/ethnicity leading to the concern that interventions or modifications to the FCI instrument intended to narrow the gender gap might not help underrepresented students of non-majority race or ethnicity.

This work was performed at one institution and the results may be dependent on the student population or the instructional environment. The work relied on self-reported race and ethnicity information and forced students to identify a single race or ethnicity which may miscount multi-race students. In the future, research will focus on the reasons for the differences between the various racial and ethnic groups.

VI. CONCLUSION

The gender gap on popular physics conceptual assessments has been thoroughly explored [3]. Race and ethnicity has been less studied. In this work, a gender gap (7.76%) was found on the FCI posttest. Differences between students by race and ethnicity were analyzed; Caucasian students outperformed both African-American and Hispanic students by 14.03% and 5.64% respectively. Controlling for course performance measured by overall physics grade, the difference between male and female students was slightly increased. Differences between students by race and ethnicity were also significant after controlling for overall physics grade, but somewhat reduced. Although main effects of gender and race/ethnicity were present in this analysis, no significant race/ethnicity by gender interaction was measured. The gender gap was shared equally by students of all races and ethnicities.

[1] D. Hestenes, M. Wells, and G. Swackhamer, Phys. Teach. **30**, 141-158 (1992).
[2] R. K. Thornton and D. R. Sokoloff, Am. J. Phys. **66**, 338-352 (1998).
[3] A. Madsen, S. B. McKagan, and E. C. Sayre, Phys. Rev. Phys. Educ. Res. **9**, 020121 (2013).
[4] L. Kost, S. Pollock, and N. Finkelstein, Phys. Rev. Phys. Educ. Res. **5**, 010101 (2009).
[5] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, Phys. Rev. Phys. Educ. Res. **6**, 020112 (2010).
[6] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Phys. Rev. Phys. Educ. Res. **3**, 010107 (2007).
[7] M. Lorenzo, C. G. Crouch, and E. Mazur, Am. J. Phys. **74**, 118-122 (2006).
[8] K. Kreutzer and A. Boudreaux, Phys. Rev. ST Phys. Educ. Res. **8**, 010120 (2012).
[9] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, AIP Conf. Proc. **1179**, 177-180 (2009).
[10] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, Science **330**, 1234-1237 (2010).
[11] R. R. Hake, Am. J. Phys. **66**, 64-74 (1998).
[12] S. Freeman, *et al.*, Proceedings of the National Academy of Sciences **111**, 8410-8415 (2014).
[13] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2013).
[14] R. Scherr, Phys. Rev. Phys. Educ. Res. *Editorial: Never mind the gap: Gender-related research in Physical Review Physics*

Education Research, 2005-2016 (2016).
[15] L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Phys. Rev. Phys. Educ. Res. **2**, 010105 (2006).
[16] Z. Hazari, R. H. Tai, and P. M. Sadler, Sci. Educ. **91**, 847-876 (2007).
[17] National Science Foundation and Nation Center for Science and Engineering Statistics, *Science and Engineering Degrees, by Race/Ethnicity of Recipients: 2002-2012. Detailed Statistical Tables NSF 15-321* (National Science Foundation, Arlington, VA, 2015).
[18] L. Perna *et al.*, Res. High. Educ. **50**, 1-23 (2009).
[19] J. Chapa and B. De La Rosa, J. Hisp. High. Educ. **5**, 203-221 (2006).
[20] V. A. Lotkowski, S. B. Robbins, and R. J. Noeth, ACT Policy Report, American College Testing ACT Inc (2004).
[21] J. M. Braxton, J. F. Milem, and A. S. Sullivan, J. High. Educ. **71**, 569-590 (2000).
[22] T. Ishitani, Res. High. Educ. **49**, 403-419 (2008).
[23] B. Toven-Lindsey, M. Levis-Firzgerald, P. H. Barber, and T. Hasson, CBE-Life Sci. Educ. **14**, ar12:1-12 (2015).
[24] D. F. Carter, New Dir. Inst. Res. **2006**, 33-46 (2006).
[25] S. Hurtado, C. B. Newman, M. C. Tran and M. J. Chang, New Dir. Inst. Res. **2010**, 5-15 (2010).