What does the Force and Motion Conceptual Evaluation pretest measure?

Dona Hewagallage and John Stewart
Department of Physics and Astronomy, West Virginia University,
135 Willey St., Morgantown, West Virginia, 26506, United States

The Force and Motion Conceptual Evaluation is commonly used to measure the conceptual understanding of Newtonian mechanics. Several studies have reported a substantial difference in pretest scores between men and women. This study examines the contribution of several prior preparation factors to explain the variance in pretest score and whether these factors explain gender differences in the pretest score. The study examined a large sample ($N = 1060$) of students taking introductory calculus-based mechanics at the university level. Women outperformed men on most prior preparation and college achievement measures. No significant differences between men and women were found in high school physics taking patterns. Linear regression analysis showed only 23% of the variance in FMCE pretest score could be explained using a linear combination of prior preparation variables. Controlling for these variables failed to explain the gender difference in pretest scores; conversely, the gender difference increased controlling for prior preparation.

2020 PERC Proceedings edited by Wolf, Bennett, and Frank; Peer-reviewed, doi.org/10.1119/perc.2020.pr.Stewart
Published by the American Association of Physics Teachers under a Creative Commons Attribution 4.0 license.
Further distribution must maintain the cover page and attribution to the article's authors.

I. INTRODUCTION

Conceptual physics pretests, tests given prior to instruction, are applied in many physics classrooms. Halloun and Hestenes used a pretest and post-test to show that little conceptual understanding was gained through traditional college physics instruction [1]. Hake collected Force Concept Inventory (FCI) [2] pretest and post-test data from multiple institutions to show that traditional instruction was generally ineffective at producing conceptual understanding [3]. In 1998, Thornton and Sokoloff introduced the Force and Motion Conceptual Evaluation (FMCE) [4] which provided somewhat less coverage of Newtonian mechanics than the FCI but addressed issues of graphical reasoning and some misconceptions more thoroughly [5, 6].

Almost since their introduction, gender differences in FCI and FMCE pretest and post-test scores have been reported with men scoring on average 13% higher on the pretest, 12% higher on the post-test [7]. The source of these differences is an active area of research with studies investigating a variety of explanations including instrumental fairness [8, 9], psychosocial factors [10, 11], and instructional pedagogy [12–17]. While there is consistent evidence that some of the gender difference in the FCI results from instrumental bias [8], no consistent explanation of gender differences in the FMCE has emerged. Henderson *et al.* showed that the items in the FMCE were generally fair to men and women [9].

Both the FMCE and the FCI have been used as a pretest in multiple studies; many of these studies investigated gender differences. In 2009, Kost *et al.* reported that gender differences in the FMCE post-test score disappear if the students are grouped based on pretest score [13]; they posited that gender differences in pretest score resulted from differences in physics preparation. They reported differences in high school physics taking between men and women, but did not use these variables in regression models. Recent studies have used pretest scores as a proxy for prior preparation in physics to understand differences between men and women in both post-test scores [18] and final exam scores [19] showing that pretest scores and other high school level preparation variables largely explain differences in final exam score, but not differences in conceptual post-test score. Henderson *et al.* speculated that some of the unexplained post-test differences might result from differences in prior preparation not captured by the pretest [18].

While some knowledge of physics is a necessary prerequisite to doing well on a pretest, it is not the only prerequisite and the origin of that knowledge is also unclear. Does the physics knowledge come from high school classes, other classes at the college level, retaking the current class, or other sources? Further, any examination measures both knowledge and general academic preparation (higher performing students tend to do better on all examinations). The purpose of this study is to first determine what factors predict pretest scores and then to determine if differences in these factors explain gender differences in pretest scores.

This study seeks to explore two research questions. *RQ1: What academic preparation and performance measures are important in predicting pretest scores? RQ2: Do differences in academic preparation and performance between men and women explain differences in pretest scores?*

II. METHODS

The FMCE [4] measures conceptual understanding of Newtonian mechanics. The test consists of 43 multiple-choice items (excluding the energy items). After its introduction, Thornton *et al.* introduced a modified scoring method that produced a total score of 33 by eliminating some items and scoring some items as groups; this method is used in the current study [5].

This study was performed from fall 2015 to spring 2019 at a large land-grant university in the eastern US. The university's general undergraduate population was 80% White, 6% international, 4% Hispanic, 4% African American, 4% students reporting two or more races, 2% Asian, and other groups each with 1% or less [20]. The ACT scores of the institution ranged from 21 to 26 (25th to 75th percentile) [20].

FMCE pretest scores ($N = 1060$) were collected in the calculus-based introductory mechanics course taken by scientists and engineers. Student demographic and college performance measures were accessed from institutional records. Student high school science course selection was collected using a survey instrument.

This work examined a broad set of high school and college-level preparation and performance variables. These variables are added as groups to regression models predicting FMCE pretest scores to form a sequence of models. Model 1 included only gender as an independent variable. General high school preparation, Model 2, is measured by high school grade point average (HSGPA), ACT or SAT mathematics percentile (ACTM), and ACT English percentile or SAT verbal percentage (ACTV). When both ACT and SAT percentile scores were available, they were averaged. General college academic success, Model 3, was measured by a student's college GPA (CGPA), the percentage of the credit hours enrolled which were completed (Cmp Pct), the total credit hours completed (Total Hrs), and the number of Science, Technology, Engineering, and Mathematics (STEM) classes completed (STEM Cls). All Model 3 variables were collected before the student enrolled in the physics class. Mathematics readiness, Model 4, was characterized by the student's mathematics entry point (MathE) representing the first mathematics class taken in college. For the vast majority of the students in the course studied, their 4-year degree plans specify taking Calculus 1 in the fall freshman semester. Students who are ready to take Calculus 1 upon arriving on campus are "math ready." Mathematics readiness is characterized by a 4-level categorical variable (MathE) with levels Algebra (Alg) indicating the first mathematics class is before calculus, Stretch Calculus (SCalc) indicating the student's first mathematics class was a

two-semester Calculus 1 class because they were not prepared for the one-semester Calculus 1, Calculus (Calc) the student first enrolled in the one-semester Calculus 1, and Advanced (Adv) the student first enrolled in a class more advanced than Calculus 1. Advanced students must have Advanced Placement (AP) or transfer credit for Calculus 1. Model 5 captured the student's high school physics preparation with a three-level categorical variable (HSPhysics) with levels no high school physics (No HS Phys), some high school physics (HS Phys), and AP high school physics taken (AP HS Phys). High school physics classes vary greatly and many students take more than one class. Multiple survey questions asked about the number and type (regular, honors, AP, dual enrollment, or International Baccalaureate). Preliminary analysis suggested the three-level HSPhysics variable captured much of the explanatory power of the more detailed responses. A student may have taken an AP physics class without taking or passing the AP examination. Model 6 measures AP credit actually earned. The variable AP physics credit (AP Phys Crd) captures whether the student received university credit for any AP physics examination. Most students who do have credit for AP physics have credit only for the algebra-based class which requires they still take the calculus-based class. Similarly AP chemistry credit (AP Chem Crd) and AP mathematics credit (AP Math Crd) capture whether the student has credit for introductory chemistry or Calculus 1 or 2.

III. RESULTS

Table I shows the descriptive statistics for the variables used in this study in aggregate and disaggregated by gender. Cohen's d characterizes the effect size for differences in the continuous variables between men and women. Cohen's criteria suggests 0.2 as a small effect, 0.5 as a medium effect [21]. The statistical significance of the differences was determined using a t-test and indicated by a superscript on d. Cramer's V was used to characterize the effect size of differences between men and women in the dichotomous or categorical variables. The criteria for V are 0.1 is a small effect with 1 degree of freedom (df), 0.07 with 2 dfs, and 0.06 with 3 dfs. The significance of the differences were determined by a chi-squared test; the significance level is represented by a superscript on V. The Cramer's V for AP chemistry and not having AP STEM credit represented small effects. For the four-level math entry variable ($df = 3$), the $V = 0.12$ is a small to medium effect (the threshold for medium effect is $V = 0.17$). Only one V value is provided for the levels of the MathE and HSPhysics categorical variables because these are single variables with multiple levels.

Table I clearly shows women in the study have a nearly universal advantage in high school and college-level performance and preparation measures. With significant advantages representing small effects in HSGPA, CGPA, ACT/SAT verbal scores, the number of STEM classes completed, the rate of completing classes, the college mathematics entry point, and the amount of AP chemistry credit. No significant differences were found in either high school physics taking patterns or earning AP physics credit.

Variable	Average Overall	Average Men	Average Women	Cohen's d
N	1060	747	313	
FMCE Pre%	23 ± 18	24 ± 18	20 ± 16	$.22^c$
FMCE Post%	46 ± 28	49 ± 28	37 ± 24	$.48^c$
HSGPA	$3.9 \pm .6$	$3.8 \pm .7$	$4.0 \pm .5$	$.37^c$
ACTM	81 ± 15	81 ± 15	82 ± 14	.08
ACTV	74 ± 18	72 ± 19	78 ± 17	$.36^c$
CGPA	$3.3 \pm .5$	$3.3 \pm .5$	$3.5 \pm .4$	$.44^c$
Cmp Pct	93 ± 12	92 ± 12	95 ± 11	$.19^b$
Total Hrs	29 ± 17	29 ± 17	29 ± 16	.03
STEM Cls	$3.8 \pm .9$	$3.7 \pm .9$	$4.0 \pm .9$	$.33^c$
	Overall %	Men %	Women %	Cramer's V
MathE Alg	29	30	27	$.12^b$
MathE SCalc	22	23	19	
MathE Calc	34	35	33	
MathE Adv	15	12	21	
No HS Phys	25	23	28	.06
HS Phys	49	50	46	
AP HS Phys	26	27	26	
No AP Cred	82	85	77	$.10^b$
AP Phys Crd	4	4	4	.00
AP Chem Crd	8	6	12	$.10^b$
AP Calc Crd	11	10	16	$.09^b$

TABLE I. Descriptive statistics. Entries represent mean $\pm$ standard deviation. Note: "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$.

All the variables in Table I could possibly influence a student's pretest score. The variables, however, are far from independent. The habits that lead to academic success in high school (HSGPA) should also lead to success in college (CGPA). A student's ACT/SAT score may be influenced by a generally enriched high school academic experience which may make access to AP STEM classes more likely. To determine the relative and combined importance of these variables they were used as independent variables in a set of linear regressions with pretest score as the dependent variable. All continuous variables were standardized by subtracting the mean and dividing by the standard deviation.

Table II shows the regression coefficients resulting from these regressions. Model 1 uses gender alone to predict pretest score showing a 0.22 standard deviation difference between men and women. Because pretest is standardized, the regression coefficients of dichotomous or categorical variables can be interpreted as the effect size (Cohen's d) change from the base level of the variable. For example, for the gender variable the regression coefficient is the number of pretest

Variable	1	2	3	4	5	6	7	8	9
*Gender(Female)	$-.16 \pm .06^b$								
Gender(Male)	$.22 \pm .07^c$								$.31 \pm .06^c$
HSGPA		–					–	–	–
ACTM		$.25 \pm .04^c$					$.13 \pm .04^b$	–	–
ACTV		$.15 \pm .04^c$					$.10 \pm .04^b$	$.13 \pm .04^c$	$.15 \pm .04^c$
ACTM x ACTV		$.11 \pm .03^c$					$.09 \pm .03^c$	$.07 \pm .03^b$	$.07 \pm .02^c$
CGPA			$.15 \pm .03^c$				$.07 \pm .03^a$	–	$.08 \pm .03^a$
CmpPct			–				–	–	–
Total Hrs			$-.13 \pm .03^c$				–	–	–
STEM Cls			–				–	$-.06 \pm .03^a$	–
*MathE Alg				$-.28 \pm .06^c$					
MathE SCalc				–			–	–	–
MathE Calc				$.48 \pm .07^c$			$.29 \pm .09^c$	$.16 \pm .08$	$.17 \pm .08^a$
MathE Adv				$.70 \pm .09^c$			$.44 \pm .10^c$	$.26 \pm .10^b$	$.30 \pm .10^b$
*No HS Phys					$-.38 \pm .06^c$				
HS Physics					$.30 \pm .07^c$			$.30 \pm .07^c$	$.29 \pm .07^c$
AP HS Phys					$.89 \pm .08^c$			$.61 \pm .08^c$	$.60 \pm .08^c$
*No AP Cred						$-.12 \pm .03^c$			
AP Phys Crd						$1.51 \pm .15^c$		$1.14 \pm .15^c$	$1.11 \pm .15^c$
AP Chem Crd						$.23 \pm .04^a$		–	–
AP Calc Crd						$.36 \pm .10^c$		–	–
R^2	$.01^b$	$.09^c$	$.05^c$	$.07^c$	$.11^c$	$.13^c$	$.11^c$	$.23^c$	$.25^c$

TABLE II. Paneled regression showing the regression coefficients with standard error. For dichotomous and categorical variables, * represents the base level of the variable and other coefficients in the group the change from this base level. A dash indicates the variable was used in the regression but was not significant. Note: "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$.

standard deviations difference between men and women. For models containing only categorical or dichotomous variables, the base level is the intercept of the model. The base level is marked with an asterisk in Table II. For example, for gender the average score of women was chosen as the base level which was 0.16 standard deviations below the sample average. The average score of men was 0.22 standard deviations above this base level or 0.06 standard deviations above the class average. The amount of the variance explained by the model is given by R^2.

Models 2 through 8 investigated the variables influencing pretest score without considering gender. Model 2 investigated general high school performance variables and explained 9% of the variance in pretest score. The interaction between ACT/SAT verbal and mathematics scores was consistently significant in all models using these variables. A one standard deviation increase in both ACT/SAT verbal and mathematics scores produced a 0.51 standard deviation increase in pretest score. Model 3 investigated college-level performance measures explaining 5% of the variance in pretest score. Only CGPA and the total credit hours completed were significant. The coefficient of the total hours was negative as was the coefficient of STEM classes completed suggesting pretest scores were not influenced by the content in the other STEM classes taken. The negative coefficient of total hours may also suggest that the time since high school physics was taken may be important.

Model 4 included only the 4-level math entry point variable explaining 7% of the variance. Being academically ready to take Calculus 1 or a more advanced class as the first math class represented a medium 0.48 to near a large 0.70 effect over the pretest scores of students who were not placed in calculus. ACT and SAT scores are used as part of the placement criteria for mathematics and, as such, Model 2 and 4 are not independent.

Model 5 investigated the role of high school physics in pretest scores. Students with no high school physics background scored 0.38 standard deviations below the class average. Having some high school physics increased pretest score by 0.30 standard deviations, a small effect. Taking an AP physics class (of any kind) increased pretest score by 0.89 standard deviations, a large effect. Students taking AP physics may not take the AP test or may not pass the test to receive AP credit.

Model 6 investigated the effect of AP STEM credit on pretest score. To receive college credit for an AP class, the student must achieve a minimum score on the AP examination. Students with AP credit for either calculus or chemistry had higher pretest scores, a small effect. Students with AP credit for physics had substantially higher pretest scores, a

very large effect. The size of this effect is incredible; even though only 4% of the students have AP physics credit, these students accounted for most of the 13% of the variance explained by Model 6. This suggests the algebra-based AP curriculum is very effective at producing conceptual understanding.

If the variables in Models 2 through 6 were independent, these variables would explain 45% of the variance in pretest score. The variables, however, are not independent. Model 7 combines models 2, 3 and 4 and contains both general high school and college preparation and performance variables. These variables together, which do not contain a measure of specific preparation in physics, explain 11% of the variance in pretest score. For models 7 and 8, only significant variables are retained in the models. Model 8 adds variables measuring the student's specific preparation in physics. These variables explained an additional 12% of the variance not explained by Model 7. The variables AP HS Phys and AP Phys Crd interact in this model to change the meaning of the AP HS Phys variable from its definition in Model 5 as whether the student had taken AP physics to its meaning in Model 8 where it indicates the student took AP physics but did not pass the AP test (a score of 3 of 5 is considered passing at the university studied). As such, taking an AP physics class increased pretest scores by 0.61 standard deviations, a medium effect; passing the AP test increased pretest score an additional 1.14 standard deviations, a large effect, correcting for general academic preparation and performance.

Two other groups of variables were tested and were not significant in all models: transfer credit and standing (freshman, sophomore, etc.). Beyond AP credit, many students also receive credit for STEM classes by transferring courses from other institutions. These may be local community college classes or university classes taken online in high school (dual enrollment classes). Having transfer credit for chemistry, calculus, or physics had no significant effect on pretest score. Each semester approximately 20% of the students in the class studied fail to complete the class successfully earning a D, F, or withdrawing. Having taken the class before had no significant effect on pretest score.

Model 9 added gender to Model 8 to determine if the gender differences in pretest score observed in Model 1 were the result of the student's prior academic preparation. Rather than reducing the differences observed in Model 1, the gender difference in pretest score, now controlling for prior general academic preparation, college academic performance, and prior preparation in physics, increased to 0.31 standard deviation, still a small effect.

IV. DISCUSSION AND CONCLUSIONS

RQ1: What academic preparation and performance measures are important in predicting pretest scores? High school physics preparation (Model 5 and AP Phys Crd) explained 17% of the variance in pretest score, while general college and high school academic preparation and performance measures (Model 7) explained 11% of the variance. All variables together explained 23% of the variance. As such, pretest scores measure a combination of general academic preparation and performance and specific preparation in physics; however, the majority (but not the overwhelming majority) of the variance is explained by high school physics class taking patterns. This study included an extensive collection of background variables, but still only explained 23% of the variance in pretest score; as such most of variance in pretest is not predictable either by academic preparation or general features of preparation in physics. This suggests pretest scores may not be an accurate characterization of a student's incoming preparation.

RQ2: Do differences in academic preparation and performance between men and women explain differences in pretest scores? Using the full set of academic variables available in this study failed to explain gender differences in pretest score. Controlling for preparation actually increased the pretest gender difference. Because women had substantially better general academic preparation (higher HSGPA, CGPA, and ACTV; all small to medium effects) and because pretest scores are partially dependent on general academic preparation and performance, the pretest scores underestimated the actual differences between men on women on the pretest.

While extensive, the variables used in this study are hardly complete. Other academic factors may be important in pretest score such as informal science experiences or high school class pedagogy. However, it seems unlikely that the missing factors are more important than those already in the model. The observation that not only are gender differences not explained but actually increased by controlling for the variables in this study makes it very unlikely that additional academic preparation or performance variables do explain a substantial part of the gender difference in pretest score. For this student population, where women were almost uniformly better prepared and higher performing, it seems likely that some other factor is the cause of the gender differences in pretest score.

The strong effect of the AP physics variable over both the transfer physics classes and non-AP high school physics classes suggest pretest scores may be very sensitive to the details of the high school physics instruction; however, it seems unlikely that these detailed differences are sufficiently unevenly distributed between men and women to explain much of the gender differences.

This work was supported by the National Science Foundation under grants ECR-1561517 and HRD-1834569.

[1] I. Halloun and D. Hestenes, Am. J. Phys. **53**, 1043 (1985).
[2] D. Hestenes, M. Wells, and G. Swackhamer, Phys. Teach. **30**, 141 (1992).
[3] R. Hake, Am. J. Phys. **66**, 64 (1998).
[4] R. Thornton and D. Sokoloff, Am. J. Phys. **66**, 338 (1998).
[5] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Phys. Rev. Phys. Educ. Res. **5**, 010105 (2009).
[6] J. Yang, C. Zabriskie, and J. Stewart, Phys. Rev. Phys. Educ. Res. **15**, 020141 (2019).
[7] A. Madsen, S. McKagan, and E. Sayre, Phys. Rev. Phys. Educ. Res. **9**, 020121 (2013).
[8] A. Traxler, R. Henderson, J. Stewart, G. Stewart, A. Papak, and R. Lindell, Phys. Rev. Phys. Educ. Res. **14**, 010103 (2018).
[9] R. Henderson, P. Miller, J. Stewart, A. Traxler, and R. Lindell, Phys. Rev. Phys. Educ. Res. **14**, 020103 (2018).
[10] L. Kost, S. Pollock, and N. Finkelstein, in *AIP conference proceedings* (AIP, 2009), vol. 1179, pp. 177–180.
[11] A. Miyake, L. Kost-Smith, N. Finkelstein, S. Pollock, G. Cohen, and T. Ito, Science **330**, 1234 (2010).
[12] M. Lorenzo, C. Crouch, and E. Mazur, Am. J. Phys. **74**, 118 (2006).
[13] L. Kost, S. Pollock, and N. Finkelstein, Phys. Rev. Phys. Educ. Res. **5**, 010101 (2009).
[14] P. Kohl and H. Kuo, in *AIP Conference Proceedings* (AIP Publishing, 2009), vol. 1179, pp. 173–176.
[15] S. Pollock, N. Finkelstein, and L. Kost, Phys. Rev. Phys. Educ. Res. **3**, 010107 (2007).
[16] M. Cahill, K. Hynes, R. Trousil, L. Brooks, M. McDaniel, M. Repice, J. Zhao, and R. Frey, Phys. Rev. Phys. Educ. Res. **10**, 020101 (2014).
[17] N. Karim, A. Maries, and C. Singh, Eur. J. Phys. **39**, 1 (2018).
[18] R. Henderson, J. Stewart, and A. Traxler, Phys. Rev. Phys. Educ. Res. **15**, 010131 (2019).
[19] S. Salehi, E. Burkholder, G. Lepage, S. Pollock, and C. Wieman, Phys. Rev. Phys. Educ. Res. **15**, 020114 (2019).
[20] *US News & World Report: Education*, US News and World Report, Washington, DC, https://premium.usnews.com/best-colleges. Accessed 2/23/2019., URL https://premium.usnews.com/best-colleges.
[21] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, NY, 1977).