

Gender & LEAP Pedagogy: What Does The Gender Force Concept Inventory Have To Say?

Twanelle Walker Majors*, Paula V. Engelhardt†

* *Department of Chemistry, Tennessee Technological University, 412A Foster Hall, Cookeville, TN 38505*

† *Department of Physics, Tennessee Technological University, 215 Bruner Hall, Cookeville, TN 38505*

Abstract: The Gender Force Concept Inventory (GFCI) was used in lieu of the Force Concept Inventory (FCI) to assess the Learner-centered Environment for Algebra-based Physics (LEAP) curriculum which utilizes interactive engagement pedagogies. Students enrolled in the 2014 spring semester PHYS2010 introductory algebra-based physics course designed for non-majors composed the sample for this phase of curriculum evaluation. The purpose of this preliminary analysis of GFCI performance was to determine if the conceptual gains of LEAP students were significantly more than traditionally taught students. Pedagogy accounted for 42% of the difference in both the posttest and the normalized gain. Gender accounted for nothing. By looking at the lack of differences between males and females when measuring with the GFCI, assessment choice becomes an interesting option for thinking about what our data has to say.

Keywords: gender, pedagogy, Gender Force Concept Inventory, non-majors

PACS: 01.40.Fk, 01.40.G-, 01.40.gb, 01.40.gf

INTRODUCTION

Differences in performance between groups have been described as the achievement gap when defined within the deficit paradigm. As a rebuttal to deficit ideology, others have criticized the deficit-focused terminology by offering critical alternatives to describe the phenomenon such as “education debt”, “the receivment gap”, and the “gap-gazing fetish” [1, 2, 3]. Deficit thinking has become the dominant discourse in the education of underrepresented or underperforming groups [4, 5]. This occurs through the use of stereotypical images to propagate deficit thinking and assumptions that education is equitable by design [5]. For the physics program at Tennessee Technological University, the deficit is proposed to be located in the ability of curriculum and pedagogy to promote conceptual understanding—not deficits within individuals.

For females, traditional pedagogies and gender-biased curricula have been linked to lower interest in physics [6, 7]. Creating female-oriented contexts for scenarios in curricular materials has been proposed as a way to increase interest and reduce negative experiences that stem from continually encountering scenarios that are unfamiliar [8]. The Force Concept Inventory [9] (FCI) has been a source of gender-bias, sometimes called gender gap, studies for a considerable amount of time [10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Laura McCullough developed the Gender Force Concept Inventory (GFCI) by altering

the gender context of each question on the FCI to a female-oriented scenario while not altering the concept or physics context [10]. The purpose was to alter question scenarios from a formal classroom context to a less formal context one might encounter in the daily routines of life [10]. The GFCI was used in lieu of the FCI to assess the Learner-centered Environment for Algebra-based Physics (LEAP) curriculum which utilizes interactive engagement pedagogies. Though results from each version of the inventory cannot be compared directly, this phase of the research probes further into ideas about assessment measures and differences between students.

LEAP is physics curriculum for algebra-based physics appropriate for both university and high school settings. The pedagogy and activity sequence builds on the work of the NSF supported project, Physics for Everyday Thinking (PET). Students work in groups to develop their understanding of various physics phenomena including forces, energy, electricity and magnetism, light and optics. Students utilize hands-on experiments and computer simulations to provide evidence to support their conceptual understanding.

This study looked at students of the spring 2014 first semester of an introductory algebra-based physics course designed for non-majors. The independent variables of interest were gender and pedagogy (LEAP versus traditional). To attempt a stronger objectivity [20], an analysis of the effects of gender and pedagogy was conducted using the posttest score as well as the normalized gain as dependent variables. The posttest

measure was an important component of this analysis, allowing ability to be controlled for thus strengthening the findings. Though normalized gain is a measure of FCI performance common to the physics education research community [21], its use in an analysis of effect can muddy the interpretation given the inclusion of the pretest in the calculation of normalized gain. However, to ensure transparency about the use of pretest scores in the analysis and thus increase objectivity, both methods are presented here.

GFCI DATA ANALYSIS

Traditional course sections typically have twice the enrollment of LEAP sections due to increased availability associated with lecture hall size. After excluding students missing the pretest or posttest, 62.7% ($n = 47$) of the sample were enrolled in the traditional section while 37.3% ($n = 28$) were enrolled in the LEAP section. For the traditional section, 70.2% ($n = 33$) were male. Of all males included in the sample, 71.7% ($n = 33$) enrolled in the traditional section while females enrolled equally in the LEAP (48.3%) and traditional (51.7%) sections.

The Shapiro-Wilk statistic indicated that the assumption of normality was met for the posttest ($p_{\text{Traditional}} = .079$, $p_{\text{LEAP}} = .331$) and for normalized gain ($p_{\text{Traditional}} = .085$, $p_{\text{LEAP}} = .638$), as was confirmed by histograms. Q-Q plots indicated the assumption of linearity of both measures was met for levels of pedagogy and for gender. Boxplots show three outliers for traditional pedagogy and none for LEAP with respect to normalized gain. Outliers were not transformed, as those values are within the range of possibility thus not considered to be a result of measurement error. The posttest outlier was from the traditional section and was not transformed since that value fell within the range of possible scores.

TABLE 1. ANCOVA summary for GFCI posttest

Source	Type III Sum of Squares	df	Mean Square	F	p	Partial η^2
Pretest	246.5	1	246.5	15.64	.00	0.183
Pedagogy	792.2	1	792.2	50.25	.00	0.418
Gender	2.71	1	2.71	0.17	.68	.002
Pedagogy x Gender	6.67	1	6.67	0.42	.52	.006
Error	1104	70	15.76			
Total	15378	75				

TABLE 2. Confidence intervals for posttest.

Group	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Traditional	10.38 ^a	.66	9.06	11.69
LEAP	17.68 ^a	.76	16.16	19.20
Male	14.26 ^a	.69	12.88	15.64
Female	13.80 ^a	.79	12.22	15.38
Traditional Male	10.29 ^a	.70	8.89	11.68
Traditional Female	10.47 ^a	1.16	8.15	12.79
LEAP Male	18.23 ^a	1.16	15.93	20.54
LEAP Female	17.13 ^a	1.03	15.08	19.19

a. Covariate evaluated at the following: Pre = 7.65

Posttest: Does Pretest Matter?

Preliminary data screening for the posttest measure showed no interaction of pedagogy and gender with the pretest, $F(3, 68) = .301$, $p = .825$, indicating that differences in posttest scores between groups were equally adjusted by pretest performance. Said another way, we may find that pretest performance explains some of the posttest variety, but differences in posttest outcomes for LEAP, traditional, male, or female will not vary as a function of the pretest. In addition, the Levene's test for equal variances indicated that equal variances can be assumed, $F(3, 71) = 1.629$, $p = .190$.

With the assumption of homogeneity of regression slopes met, a univariate analysis of covariance (ANCOVA) was conducted on posttest score while controlling for pretest score. The ANCOVA summary in Table 1 indicated that there was no interaction between pedagogy and gender, improving the interpretability of the main effects of the independent variables and the covariate. The ANCOVA revealed a main effect of pretest, $F(1, 70) = 15.635$, $p = .000$, partial $\eta^2 = .183$. Score on the pre-test significantly adjusted the posttest and accounts for 18.3% of the variance in post-test scores. The ANCOVA failed to reveal a main effect of gender on the post-test, $F(1, 70) = .172$, $p = .680$, partial $\eta^2 = .002$. There was no significant difference in posttest scores for males and females. After adjusting for the covariance of pretest, there is sufficient evidence that there is a difference between means of those receiving LEAP pedagogy and those who did not, $F(1, 70) = 50.251$, $p = .000$, partial $\eta^2 = .418$, so that pedagogy is found to be associated with 41.8% of the variance in posttest performance.

Table 2 shows that the 95% confidence intervals of traditional and LEAP pedagogies did not overlap, further explaining the significance of the difference in posttest performance between two pedagogical approaches. This is a striking statistic when one considers that the lower bound of the mean of LEAP students is statistically significantly higher than the upper bound of the mean of traditional students. For gender, confidence intervals show nearly complete overlap of the posttest mean of females with the posttest mean of males. A similar overlap is found for the interaction of pedagogy and gender, further explaining the insignificance of differences in the posttest means of subgroups. Having high confidence in this finding is especially important given the concern that LEAP pedagogy is more efficacious for some student groups and that traditional pedagogies may be more detrimental for some student groups.

Normalized Gain

Levene's Test indicated that the assumption of equality of variance in normalized gains was met, $F(3, 71) = 1.192, p = .319$. Univariate analysis of variance (ANOVA) was then conducted to determine if differences in normalized gains were significant and if there was an overlap in the portions of variance explained by gender and pedagogy. Looking at covariance with pretest score was not appropriate, as main effects would not be interpretable due to the use of the pretest in the calculation of the normalized gain.

The ANOVA summary in Table 3 showed no interaction between pedagogy and gender, which improved the interpretability of the effects of pedagogy and gender. Since the interaction would only explain 2.1% of the variance in normalized gains, looking further at the insignificance of this interaction seemed moot given the high insignificance of gender. A look at Table 4 showed the near complete overlap of 95% confidence intervals for normalized gain means of traditional males and females as well as LEAP males and females which supported that decision.

TABLE 3. ANOVA summary for GFCI normalized gain

Source	Type III Sum of Squares	df	Mean Square	F	<i>p</i>	Partial η^2
Pedagogy	1.839	1	1.839	52.08	.00	.423
Gender	.008	1	.008	.21	.65	.003
Pedagogy x Gender	.055	1	.055	1.55	.22	.021
Error	2.507	71	.035			
Total	9.108	75				

TABLE 4. Confidence intervals for normalized gain.

Group	Mean	Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Traditional	.126	.030	.066	.185
LEAP	.461	.036	.390	.532
Male	.304	.031	.243	.366
Female	.283	.035	.213	.352
Traditional Male	.107	.033	.042	.173
Traditional Female	.144	.050	.044	.244
LEAP Male	.501	.052	.397	.605
LEAP Female	.422	.049	.325	.518

The ANOVA failed to reveal a main effect of gender on normalized gain, $F(1, 71) = .213, p = .646$, partial $\eta^2 = .003$, meaning that there are no differences in normalized gain between males and females. The lower and upper bounds of normalized gain means for males and females also explain the large insignificance of gender. There is sufficient evidence that there is a difference between normalized gain means of those receiving LEAP pedagogy and those who did not, $F(1, 71) = 52.081, p = .000$, partial $\eta^2 = .423$, so that pedagogy is found to be associated with 42.3% of the variance in normalized gains. The lack of overlap in 95% confidence intervals of normalized gain means for traditional and LEAP students further explains the significance of the contribution of LEAP pedagogy in explaining the differences in this sample.

Pre-Post Versus Normalized Gain

The effect of pedagogy on GFCI performance was measured by posttest scores as well as by normalized gains. Table 5 illustrates the outcomes of each method. Both methods of analysis showed that pedagogy was a significant variable in explaining performance on the GFCI, accounting for 42% of the differences in student performance.

TABLE 5. Outcomes comparison by method.

Variable	Method	<i>p</i>	Partial η^2	Adjusted R^2
Pedagogy	Normalized Gain	.000	.423	.423
	Posttest	.000	.418	.547
Pretest	Posttest	.000	.183	.547

CONCLUDING THOUGHTS

Prior preliminary evaluations of the LEAP program involved the use of the FCI for measuring conceptual change. For the preliminary analysis of normalized gains, ANOVA showed significantly higher gains for LEAP students, $F(1, 742) = 196.258$, $p = .000$, partial $\eta^2 = .209$, with pedagogy being associated with 20.9% of performance on the FCI. There was no interaction between pedagogy and gender. Though gender was insignificant and explained only 0.9% of the differences in FCI performance, $F(1, 742) = 6.615$, $p = .010$, partial $\eta^2 = .009$, the 95% confidence intervals for males and females in the traditional sections barely overlapped while intervals for males and females of the LEAP sections overlapped well. This questioned the idea that male and female performance is not different when you control for ability. When looking at differences between students within a semester, lower gains for females were sometimes found. Curious to look at elements contributing to differences, gender-bias of the FCI was thought worthy of consideration.

In the spring 2014 PHYS2010 course, the GFCI was used in lieu of the FCI to assess the LEAP curriculum. The findings presented here indicate that, when measured using the GFCI, pedagogy accounted for 42% of the sample differences. Though caution should be taken in comparing forms of the FCI or in comparing a large sample to this particular semester of students, there is a level of curiosity surrounding the fact that switching assessments coincided with twice the explained variance. Propensity scoring methods are planned and may illuminate this further.

For any instructional intervention development, the target goal may be a pedagogy that correlates to significant learning by students without regard to ability. In that sort of paradigm, demographics once regarded as contexts for intervention and thus differences to be “overcome” may be redefined as multiple contexts for learning. Further troubling the issues, a strong look at context within a measurement instrument would pair well with the prevalence of study of context in curricular materials. Without this stronger objectivity [20], are we communicating that we do not think it changes the data by which we compare people? This leads me to thoughts on the pervasive public perception of performance of females in STEM coursework, majors, and professions. If nontraditional pedagogy correlates no differently to the performance of females than males but does correlate to an increase in performance for all, then gender differences can be more appropriately addressed by some mode of thinking other than a deficit model.

ACKNOWLEDGMENTS

We thank Dr. George Chitiyo for critical review of the interpretation of these findings. This material is based upon work supported by the National Science Foundation under grants No. 0737324 and 1245684.

REFERENCES

1. G. Ladson-Billings, *Educ. Res.* **35**(7), 3–12 (2006).
2. T. Chambers, *J. Negro Educ.* **78**(4), 417–431 (2009).
3. R. Gutiérrez, *J. Res. Math. Educ.* **39**(4), 357–364 (2008).
4. P. Gorski, *Educ. Leadership*, **65**(7), 32–36 (2008).
5. P. Gorski, “Unlearning deficit ideology and the scornful gaze: Thoughts on authenticating the class discourse in education,” In *Assault on kids: How hyper-accountability, corporatization, deficit ideology, and Ruby Payne are destroying our schools*, edited by R. Ahlquist, P. Gorski, & T. Montaña, (Peter Lang, New York, 2011), pp. 152–175.
6. L. Hoffman, *Learn. Instr.* **12**(4), 447–465 (2002).
7. A. Zohar & B. Bronshtein, *Int. J. Sci. Educ.* **27**(1), 61–77 (2005).
8. A. Baram-Tsabari & A. Yarden, *Res. Sci. Technol. Educ.* **26**(1), 75–92 (2008).
9. D. Hestenes, M. Wells, & G. Swackhamer, *Phys. Teach.* **30**(3), 141–151 (1992).
10. L. McCullough, & T. Foster, “A gender context for the Force Concept Inventory,” 2001 AAPT Winter Meeting, San Diego, CA.
11. L. McCullough, “Gender, Math, and the FCI,” 2002 Physics Education Research Conference, Boise, Idaho.
12. M. Dancy, “The myth of gender neutrality,” *AIP Conf. Proc.* **720**, 31 (2004).
13. L. McCullough, *J. Int. Womens Stud.* **5**(4), 20–30 (2004).
14. M. Lorenzo, C. Crouch, & E. Mazur, *Am. J. Phys.* **74**(2), 118–122 (2006).
15. A. Noack, T. Antimirova, M. Milner-Bolotin, *Can. J. Phys.* **87**(12), 1269–1274 (2009).
16. L. Kost-Smith, Ph.D. dissertation, University of Colorado Boulder, 2011.
17. V. Coletta, J. Phillips, & J. Steinert, “FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects” *AIP Conf. Proc.* **1413**, 23 (2012).
18. R. Dietz, R. Pearson, M. Semak, & C. Willis, “Gender bias in the force concept inventory?,” *AIP Conf. Proc.* **1413**, 171 (2012).
19. J. Blue & P. Heller, “Using matched samples to look for sex differences,” *PERC Conf. Proc.*, **720**, 45–48 (2003).
20. S. Harding, in *Feminist Epistemologies*, edited by L. Alcoff & E. Potter (Routledge, New York, 1993), pp. 49–82.
21. R. Hake, *Am. J. Phys.* **66**(1), 64–74 (1998).